

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2005, 01:35:43 ; Search time 7976 Seconds  
(without alignments)  
11002.056 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtgcctgcggccctcgcgcg.....cggcgctgattgcgaacctcg 1811

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607355 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ha:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	1811	6	AX077432 Sequence
2	1811	100.0	1811	9	AK075358 Homo sapi
3	1631.4	90.1	1839	9	AY007158 Homo sapi
4	1621	89.5	1775	6	BD158346 Primer fo
5	1621	89.5	1775	6	AX880378 Sequence
6	1621	89.5	1775	9	AK027733 Homo sapi
7	1597.6	88.2	1709	9	AK123762 Homo sapi
8	1434.8	79.2	1614	6	BD158376 Primer fo
9	1434.8	79.2	1614	6	AX880429 Sequence
10	1434.8	79.2	1614	6	AK027741 Homo sapi
11	1371	75.7	1407	9	BC022092 Homo sapi
12	854	47.2	199522	9	AC089999 Homo sapi
13	661.4	36.5	784	6	BD149092 Primer fo
14	661.4	36.5	784	6	AX869030 Sequence
15	562	31.0	597	6	BD149164 Primer fo
16	562	31.0	597	6	AX869102 Sequence
17	550	30.4	816	6	AX781130 Sequence
18	494	27.3	561	6	BD154159 Primer fo
19	494	27.3	561	6	AX874097 Sequence

20	472.8	26.1	1485	10	BC021365	BC021365 Mus muscu
21	465	25.7	602	6	CO729207	CO729207 Sequence
22	369	20.4	572	6	AX779761	AX779761 Sequence
23	353.8	19.5	263827	2	AC094579	AC094579 Rattus no
24	353.8	19.5	274026	2	AC129633	AC129633 Rattus no
25	341.8	18.9	578	6	BD154095	BD154095 Primer fo
26	341.8	18.9	578	6	AX874033	AX874033 Sequence
27	299.2	16.5	237928	2	AC137208	AC137208 Rattus no
28	273.2	15.1	221622	10	AC125183	AC125183 Mus muscu
29	241	13.3	199165	2	AC131195	AC131195 Rattus no
30	186.2	10.3	60319	2	AC102037	AC102037 Mus muscu
31	130.8	7.2	292	6	AX315440	AX315440 Sequence
32	126	7.0	60319	2	AC102037	AC102037 Mus muscu
33	125	6.9	244529	2	AC121112	AC121112 Mus muscu
34	82.8	4.6	125020	9	AP429315	AP429315 Homo sapi
35	61.6	3.4	125020	9	AP429315	AP429315 Homo sapi
36	60	3.3	60	6	CO535257	CO535257 Sequence
37	57.6	3.2	3153	5	BC070862	BC070862 Xenopus 1
38	56.8	3.1	174471	2	AC147859	AC147859 Macropus
39	56.6	3.1	141657	2	AC147760	AC147760 Macropus
40	56.6	3.1	163688	2	AC148274	AC148274 Macropus
41	55.2	3.0	1365	5	CR385236	CR385236 Gallus ga
42	54.4	3.0	24250	9	AY764389	AY764389 Homo sapi
43	54.4	3.0	39970	9	AP001167	AP001167 Homo sapi
44	54.4	3.0	146451	2	AC019309	AC019309 Homo sapi
45	54.4	3.0	185984	9	AP000942	AP000942 Homo sapi

#### ALIGNMENTS

RESULT 1  
LOCUS AX077432 1811 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 3 from Patent WO0107607.  
ACCESSION AX077432  
VERSION AX077432.1 GI:13121982  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 Ota,T., Iwaga,T., Nishikawa,T. and Kawai,Y.  
Full length cDNA clones and proteins encoded thereby  
JOURNAL Patent: WO 0107607-A 3 01-FEB-2001;  
Helix Research Institute (JP)

FEATURES  
source location/Qualifiers  
1..1811  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
659..1468  
/note="unnamed protein product"

#### CDS

/codon\_start=1  
/protein\_id="CAC32398.1"  
/db\_xref="GI:13121983"  
/translation="MKTPELVASGMQTLQHRCRGGYRVKARTSYVDLTFGSPAG  
TRPTPDPEPWEKANKTRGVGEASALGAKSCETTPARGSTPLTPKKNYRP  
ISHTPSYCDESLFGSRSEGAERGAAPMAKGAALRLALMTWPPPTPRGSHGPRPEAP  
LRATHPAGSTPEGPAAADSOXLMGGLHSRRLKRGLSHLTLINVSQTHPRAISAP  
HTNGPQDLRSTSGVTFRRSLVTRASVSISVSTPRRGATQKPRPMK"

#### ORIGIN

Query Match 100.0%; Score 1811; DB 6; Length 1811;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTGCCTGCGGCGCTTGGCGGCTTGGTACACACGACCTGAGTGGACACAGAGAC 60  
DB 1 AGTGCCTGCGGCGCTTGGCGGCTTGGTACACACGACCTGAGTGGACACAGAGAC 60  
QY 61 CCTTCATGTTTAGAGGACCTCTGGGCTTCAAGAGCTGACGCCCGCCCTGGGCGGAC 120

```

Db      61  CTTCTCATGTTTAAAGGACCTCTGGGCTCAGGAGGTGGCGCGCCCTGGGCGGAC 120
Qy      121  TCCCCCATCCGGGGGCGGAAATGTCGGGGTGGCGTCCGAGTGTCTGCTGCTCC 180
Db      121  TCCCCCATCCGGGGGCGGAAATGTCGGGGTGGCGTCCGAGTGTCTGCTGCTCC 180
Qy      181  TGGTGTGCGGGTGAAGTGTGGGGTCTGGGGTTCTGGAATGGCGGGCGGTTCAAGCT 240
Db      181  TGGTGTGCGGGTGAAGTGTGGGGTCTGGGGTTCTGGAATGGCGGGCGGTTCAAGCT 240
Qy      241  AGCTGTGCGGGTCTCTGGGGTGAAGTCCGCGCGCGGGTCCCGGAGCGGCTTAGG 300
Db      241  AGCTGTGCGGGTCTCTGGGGTGAAGTCCGCGCGCGGGTCCCGGAGCGGCTTAGG 300
Qy      301  TGGCGGGGGTCCGGGGGCGGAGGATTCGGGGTGGCAAGATTGAACGGGGATCCG 360
Db      301  TGGCGGGGGTCCGGGGGCGGAGGATTCGGGGTGGCAAGATTGAACGGGGATCCG 360
Qy      361  CCGGGGGCGCGCGCGCTCAAGCGGGTCCAGACCTGGTGGGGAAGAAGTGGGGAC 420
Db      361  CCGGGGGCGCGCGCGCTCAAGCGGGTCCAGACCTGGTGGGGAAGAAGTGGGGAC 420
Qy      421  GGGTCCCTGAGATCCCGATGCTTACGAGCGCAAGATGCTGAGCTTTATAGGTGAC 480
Db      421  GGGTCCCTGAGATCCCGATGCTTACGAGCGCAAGATGCTGAGCTTTATAGGTGAC 480
Qy      481  CACATGTGACTTCACTCAAGTGTGATCCGTAATAATGCAAAATTCGAAGCTACTT 540
Db      481  CACATGTGACTTCACTCAAGTGTGATCCGTAATAATGCAAAATTCGAAGCTACTT 540
Qy      541  CAGTGTGTTGAGAGATTAAATGAACAATGCTTGAAGAAGCTTTTGAAGAGAGG 600
Db      541  CAGTGTGTTGAGAGATTAAATGAACAATGCTTGAAGAAGCTTTTGAAGAGAGG 600
Qy      601  TCGGAACAGAGGCTTGGCGCGGAGAGACACTGCTGTCAACAGGAGCAACAGCAG 660
Db      601  TCGGAACAGAGGCTTGGCGCGGAGAGACACTGCTGTCAACAGGAGCAACAGCAG 660
Qy      661  GAAAGACCCCGTGGAGTGGCGCTGATGGGATGACAACTCCGCTTCAAGCACTG 720
Db      661  GAAAGACCCCGTGGAGTGGCGCTGATGGGATGACAACTCCGCTTCAAGCACTG 720
Qy      721  CCGAGGTGCTAACCGGGTCAAGGCGCAAGAGTCAATATGGAGTGAAGCTCTGTT 780
Db      721  CCGAGGTGCTAACCGGGTCAAGGCGCAAGAGTCAATATGGAGTGAAGCTCTGTT 780
Qy      781  CCGAGGAGGACCCGCTTACCCCAACCGGACTTTCGATCCGCTGGGTGAGAGGCT 840
Db      781  CCGAGGAGGACCCGCTTACCCCAACCGGACTTTCGATCCGCTGGGTGAGAGGCT 840
Qy      841  CAGAACAGAGGCTGGGCAAGAGGATGAGAGGCTTGGGGCAAGGGAGCTGTGA 900
Db      841  CAGAACAGAGGCTGGGCAAGAGGATGAGAGGCTTGGGGCAAGGGAGCTGTGA 900
Qy      901  GACCAACCCCTCAAGGGGAGCAACCCCACTTCACACCAAGGAAGAACAATACAG 960
Db      901  GACCAACCCCTCAAGGGGAGCAACCCCACTTCACACCAAGGAAGAACAATACAG 960
Qy      961  ACCCATGAGCACAACCCGCTTACTGTGATGAGTCCGCTGTTGGCTCCGATGGA 1020
Db      961  ACCCATGAGCACAACCCGCTTACTGTGATGAGTCCGCTGTTGGCTCCGATGGA 1020
Qy      1021  CGCCAGCTTGGGGGCGCGGGGATGCGAAGGGGATGCGCAAGCTCGTGTCTCT 1080
Db      1021  CGCCAGCTTGGGGGCGCGGGGATGCGAAGGGGATGCGCAAGCTCGTGTCTCT 1080
Qy      1081  GTGAGCCCAACCACTTACCCCAAGGGGATGAGCACTGCGCCCGCCAGGGAGCA 1140
Db      1081  GTGAGCCCAACCACTTACCCCAAGGGGATGAGCACTGCGCCCGCCAGGGAGCA 1140
Qy      1141  GCGAGCATTCACCACTGCTGCTCTCAAGAGACCGGGGCGACAGCTCCCA 1200

```

```

Db      1141  GCGAGCATTCACCACTGCTGCTCTCAAGAGACCGGGGCGACGGGAGACTCCCA 1200
Qy      1201  GAAGTTATCTAATGGGTGGTTTACACTTTCACGCCCCCTGAAGGGGAGCTTCC 1260
Db      1201  GAAGTTATCTAATGGGTGGTTTACACTTTCACGCCCCCTGAAGGGGAGCTTCC 1260
Qy      1261  CTTACACCACTGAATGTCGCCAGCACTGGTCAATCCAGCCACCAAGTCCGCCA 1320
Db      1261  CTTACACCACTGAATGTCGCCAGCACTGGTCAATCCAGCCACCAAGTCCGCCA 1320
Qy      1321  TGGGCTCAGAGATCTCAAGGCTTTCAGTCAAGGGTGAAGCTTCCGAGCCCTG 1380
Db      1321  TGGGCTCAGAGATCTCAAGGCTTTCAGTCAAGGGTGAAGCTTCCGAGCCCTG 1380
Qy      1381  TTTCCAGGGCTGCTGAGTTAGCAATTCAGTCCATCAACCCAGAGAGTGGGGC 1440
Db      1381  TTTCCAGGGCTGCTGAGTTAGCAATTCAGTCCATCAACCCAGAGAGTGGGGC 1440
Qy      1441  CCAAGAACCAAGCCCTTGAATGATCTTTCATCAGAGGTTGCTATGGGGCCAC 1500
Db      1441  CCAAGAACCAAGCCCTTGAATGATCTTTCATCAGAGGTTGCTATGGGGCCAC 1500
Qy      1501  GGGGACAGTATGAGGCTTTCAGAGGATGAGAGCAATTCATCAGAGGAAACCC 1560
Db      1501  GGGGACAGTATGAGGCTTTCAGAGGATGAGAGCAATTCATCAGAGGAAACCC 1560
Qy      1561  TATTAAAGAGCCCTGTTGGGGGAGAGACATACAGAGGGGAGGAGGCTCTTT 1620
Db      1561  TATTAAAGAGCCCTGTTGGGGGAGAGACATACAGAGGGGAGGAGGCTCTTT 1620
Qy      1621  ATCTGACATCTCTAGTGAATCTTTCCTTTCCTCCGATTCGAGATTTGGGGG 1680
Db      1621  ATCTGACATCTCTAGTGAATCTTTCCTTTCCTCCGATTCGAGATTTGGGGG 1680
Qy      1681  CTCTAAGATGCTCTCTCCAGGCTTCTCAACCAATCTCAATTAATGTCACCA 1740
Db      1681  CTCTAAGATGCTCTCTCCAGGCTTCTCAACCAATCTCAATTAATGTCACCA 1740
Qy      1741  GGGCTGGACCTTCCACATCATTCATTTCTGTCGCAAGTGGGAGGCTGAT 1800
Db      1741  GGGCTGGACCTTCCACATCATTCATTTCTGTCGCAAGTGGGAGGCTGAT 1800
Qy      1801  TGCCAACTGG 1811
Db      1801  TGCCAACTGG 1811

RESULT 2
AK075358
LOCUS      1811 bp      mRNA      linear      PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA PSEC0043 fls, clone NT2RP1001002.
ACCESSION AK075358
VERSION   AK075358.1 GI:22761393
KEYWORDS  oligo capping; fls (full insert sequence) .
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Oca,T., Nishikawa,T., Suzuki,Y., Kawai-Hio,Y., Hayashi,K.,
Ishii,S., Saito,K., Yamamoto,J., Nakamura,A., Nagai,T.,
Nakamura,Y., Nagahari,K., Sugano,S. and Isogai,T.
HRI human cDNA sequencing project
Unpublished
2 (bases 1 to 1811)
REFERENCE Isogai,T. and Yamamoto,J.
TITLE     HRI human cDNA sequencing project
JOURNAL   Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kiserazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);

```

cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center.

## FEATURES

source

1..1811

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="NT2RP1001002"

/cell\_line="NT2"

/cell\_type="teratocarcinoma"

/clone\_1lb="NT2RP1"

/note="cloning vector: pUC19FL3-mRNA from NT2 neuronal precursor cells after 48-hours retinoic acid (RA) induction"

559..1468

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAC11568.1"

/db\_xref="gi:22761394"

/translation="MKTPVELAVSGMOTLGHORCRGGYRVKARTSYVDLTFSGPAGTRPTPDPDPVMEKANTRGVKEASAKALAKSGSCETTPRGSGPTLTPKKNKVRPISHTPSYDESLFGRSEGAFTGAPRMAKGDAAKRALIMTPPTPRSHSRPREAPLRAIHAPSPSTKEPAPADSQLSMGALSHSPKRGLSHLTILNVSTGHPATSAPTHNGPQDLRPSITSGVTFPSPLVTSRARSVISVPSPTPRRGATOKPKPMK"

## ORIGIN

Query Match 100.0%; Score 1811; DB 9; Length 1811;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGTGCCTGCGGCGCTCGGCGGCTCTAGTACACAGCACTGAGTGAAGGACCAAGAGAC 60
DB 1 AGTGCCTGCGGCGCTCGGCGGCTCTAGTACACAGCACTGAGTGAAGGACCAAGAGAC 60
QY 61 CCTCTCAATGTTTAAAGAACCTCTCTGAGCTCTAGAGCGTGCCTGCGGCGGAGAC 120
DB 61 CCTCTCAATGTTTAAAGAACCTCTCTGAGCTCTAGAGCGTGCCTGCGGCGGAGAC 120
QY 121 TCCCCCATCGCGGCGGCGGAGTGGTCCGCGGCGGCTCCGAGTCTCTGCTGCTCC 180
DB 121 TCCCCCATCGCGGCGGCGGAGTGGTCCGCGGCGGCTCCGAGTCTCTGCTGCTCC 180
QY 181 TGGTTGCTGGGTGCAAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 181 TGGTTGCTGGGTGCAAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 241 AGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
DB 241 AGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 301 TGGCGGGGGTCCGCGGCGGCGGAGCTTCGCGGCTGCAATTGACCGGGATGCA 360
DB 301 TGGCGGGGGTCCGCGGCGGCGGAGCTTCGCGGCTGCAATTGACCGGGATGCA 360
QY 361 CCGCGGCGGCGGCGGCGGCTCTACCGAGGAGTCCGAGTGGGAGAAAGGTGCGGGAG 420
DB 361 CCGCGGCGGCGGCGGCGGCTCTACCGAGGAGTCCGAGTGGGAGAAAGGTGCGGGAG 420
QY 421 GGGTCCCTGAGAGATCCGATGCTTACGAGCAAGATGCTTATAGTGTGACCTTA 480
DB 421 GGGTCCCTGAGAGATCCGATGCTTACGAGCAAGATGCTTATAGTGTGACCTTA 480
QY 481 CACATGTGACTTTCACCTCAAGTTTGTGATCCGTAATAATGACAAATTGAAAGCTTCA 540
DB 481 CACATGTGACTTTCACCTCAAGTTTGTGATCCGTAATAATGACAAATTGAAAGCTTCA 540
QY 541 CAGTGTCTTGAAGAGATTAATGAACAATGCTTGAAGCTCTTTGACAGAGGAGAC 600
DB 541 CAGTGTCTTGAAGAGATTAATGAACAATGCTTGAAGCTCTTTGACAGAGGAGAC 600
QY 601 TCGGAAGAGAGGCGCTGCGCGGAGAGCACTGCTGTACCAAGGAGCAACAGGAGCAT 660
DB 601 TCGGAAGAGAGGCGCTGCGCGGAGAGCACTGCTGTACCAAGGAGCAACAGGAGCAT 660
```

```
QY 661 GAAGACCCCGTGAAGTGTGCGGCTGAGTGAAGATGAGACCTCGGCTTTCAGACCGCTG 720
DB 661 GAAGACCCCGTGAAGTGTGCGGCTGAGTGAAGATGAGACCTCGGCTTTCAGACCGCTG 720
QY 721 CCGAGTGTGCTAACCGGGCTCAAGGCGGAGAGCTCATATGTGATGAGACTCTGTTGGCAG 780
DB 721 CCGAGTGTGCTAACCGGGCTCAAGGCGGAGAGCTCATATGTGATGAGACTCTGTTGGCAG 780
QY 781 CCCAGACGACCCCGGCTTACCCCAACCGGACTTGATTCGCGCTGCGGTGAGAAAGCTTA 840
DB 781 CCCAGACGACCCCGGCTTACCCCAACCGGACTTGATTCGCGGTGAGAAAGCTTA 840
QY 841 CAGAACGAGAGCGGTGAGCAAGAGAGCATTCGAGGCTTGGGGGCAAGGGAGCTGTA 900
DB 841 CAGAACGAGAGCGGTGAGCAAGAGAGCATTCGAGGCTTGGGGGCAAGGGAGCTGTA 900
QY 901 GACCAACCCCTCAAGGGGCAAGACCCCAACCTTCAACCAAGAAAGAAACAAATACAG 960
DB 901 GACCAACCCCTCAAGGGGCAAGACCCCAACCTTCAACCAAGAAAGAAACAAATACAG 960
QY 961 ACCCATCAAGCAACACCCGCTTACTGTGATGATGCTGCTGTTGGCTCCGATCTGAAG 1020
DB 961 ACCCATCAAGCAACACCCGCTTACTGTGATGATGCTGCTGTTGGCTCCGATCTGAAG 1020
QY 1021 CGCCAGCTTGGGGGCGCGGAGTGGGAGAGGGGATGCGGCAAAAGCTCGGTGCTCTT 1080
DB 1021 CGCCAGCTTGGGGGCGCGGAGTGGGAGAGGGGATGCGGCAAAAGCTCGGTGCTCTT 1080
QY 1081 GTGACGCGCAACCACTACCCCAAGGGGAGCACTCGCGCGGCGGCGGAGAGCACT 1140
DB 1081 GTGACGCGCAACCACTACCCCAAGGGGAGCACTCGCGCGGCGGCGGAGAGCACT 1140
QY 1141 GCGAGCAATTCACCAAGTGTGCTCTTCCCAAGACAGAGCGGGGCGACGCGGAGACTCCA 1200
DB 1141 GCGAGCAATTCACCAAGTGTGCTCTTCCCAAGAGAGAGCGGGGCGACGCGGAGACTCCA 1200
QY 1201 GAGGTTATCTAATGGGTGGGTATACACTCTTCAAGCGCGCGCGGAGAGGGGGACTTCCCATTC 1260
DB 1201 GAGGTTATCTAATGGGTGGGTATACACTCTTCAAGCGCGCGCGGAGAGGGGGACTTCCCATTC 1260
QY 1261 CCTCAACCACTGATATGCTCCCAAGCACTGATTCAGAGCAACAGTGCCTCCCAACAA 1320
DB 1261 CCTCAACCACTGATATGCTCCCAAGCACTGATTCAGAGCAACAGTGCCTCCCAACAA 1320
QY 1321 TGGGCTCTGAGATCTCAGGCTTTCACGTCAGGGGATGACTTCCGAGCGCTTGTGAC 1380
DB 1321 TGGGCTCTGAGATCTCAGGCTTTCACGTCAGGGGATGACTTCCGAGCGCTTGTGAC 1380
QY 1381 TTCCAGGGCTGCTCAGTGTGACATTTAGTGCCTTACCCCAAGAGAGAGTGGGCGAC 1440
DB 1381 TTCCAGGGCTGCTCAGTGTGACATTTAGTGCCTTACCCCAAGAGAGAGTGGGCGAC 1440
QY 1441 CCAGAAACCAAGGCGCGCTTGAATGATCTTCTTCAACAGGGTGTGCTATGAGGGCCAC 1500
DB 1441 CCAGAAACCAAGGCGCGCTTGAATGATCTTCTTCAACAGGGTGTGCTATGAGGGCCAC 1500
QY 1501 GCGGACAGATATGAGCGCTTTCAGAGGTGAGAGCAATTCACCAAGGAAACCCAGG 1560
DB 1501 GCGGACAGATATGAGCGCTTTCAGAGGTGAGAGCAATTCACCAAGGAAACCCAGG 1560
QY 1561 TATTTAAGAAAGCCCTGCTGAGGGGAGAGACATGACAGAGGGGTGGGAGTGCCTCTT 1620
DB 1561 TATTTAAGAAAGCCCTGCTGAGGGGAGAGACATGACAGAGGGGTGGGAGTGCCTCTT 1620
QY 1621 ATCTGACAAATCTGATGATCTGCTGCTTCTTCTCCGATTCGGAATTTGGGGGCGAC 1680
DB 1621 ATCTGACAAATCTGATGATCTGCTGCTTCTTCTCCGATTCGGAATTTGGGGGCGAC 1680
QY 1681 CTCTAAGATGCTCTCTCAGCGCTGTCTCAACCATATCTCAAAATTAATGTCACCCAGG 1740
DB 1681 CTCTAAGATGCTCTCTCAGCGCTGTCTCAACCATATCTCAAAATTAATGTCACCCAGG 1740
```

1741 GGCGTGGACCTCCACATCATTCATGTGCTGGCAAGTGGAATTAACGGCGTAT 1800  
1741 GGCGTGGACCTCCACATCATTCATGTGCTGGCAAGTGGAATTAACGGCGTAT 1800  
1801 TGCCAACTGG 1811  
1801 TGCCAACTGG 1811

RESULT 3  
AY007158 1839 bp mRNA linear PRI 31-AUG-2000  
LOCUS Homo sapiens clone CDABP013 mRNA sequence.  
DEFINITION AY007158  
ACCESSION AY007158.1 GI:9956071  
VERSION F11.CDNA.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1839)  
AUTHORS Anderson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.  
TITLE A 'double adaptor' method for improved shotgun library construction  
JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)  
MEDLINE 96207227  
PUBMED 8619474

REFERENCE 2 (bases 1 to 1839)  
AUTHORS Yu,W., Anderson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W., Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.  
TITLE Large-scale concatenation cDNA sequencing  
JOURNAL Genome Res. 7 (4), 353-358 (1997)  
MEDLINE 97264341  
PUBMED 9110174

REFERENCE 3 (bases 1 to 1839)  
AUTHORS Zhou,J., Yu,W., Tang,H., Mei,G., Tsang,Y.T.M., Bouck,J., Gibbs,R.A. and Margolin,J.F.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-2000) Human Genome Sequencing Center and Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA  
COMMENT The clone request should be directed to Dr. J. Margolin at Pediatrics-Hematology & Oncology, Texas Children's PEIGIN Center 10251, Houston, Texas 77030, USA. Telephone: 713-770-4583  
MargoIinbcm.tmc.edu.

FEATURES  
source  
Location/Qualifiers  
1..1839  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CDABP013"  
/clone\_lib="constructed by Y.T.M. Tsang"  
/dev\_stage="Infant"  
/note="from patient with acute lymphoblastic leukemia"

ORIGIN  
Query Match 90.1%; Score 1631.4; DB 9; Length 1839;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 163; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

179 CTTGGTGGCTGGGCAAGGCTGAGTCTGAGTTCTGATTCGAGGCGCTTTCACAC 238  
1 CTTGGTGGCTGGGCAAGGCTGAGTCTGAGTTCTGATTCGAGGCGCTTTCACAC 60  
239 GTAGCCTGTGGCGGCTCTCTGGGAGTCCGTCGCGCGCGGCTGCCCGGAGCGCTTAG 298  
61 GTAGCCTGTGGCGGCTCTCTGGGAGTCCGTCGCGCGCGGCTGCCCGGAGCGCTTAG 120  
299 GCTGGCGGGGCTCCGGGGCCCAAGGCAATCCGGGCTGAGATTGAAGGGGATCCCGGATG 358  
121 GCTGGCGGGGCTCCGGGGCCCAAGGCAATCCGGGCTGAGATTGAAGGGGATCCCGGATG 180  
359 CACGCGCGCGCGCGCGCTCAACGAGGCTCAGACTGTGGGAAAGAGTGGCGGG 418

181 CACGCGCGCGCGCGCGCTCAACGAGGCTCAGACTGTGGGAAAGAGTGGCGGG 240  
419 AGCGGTCCTCGAGAGATCCCGATGCCCTACAGAGCCAAATGCTTAATAGGTGACC 478  
241 ACGGGTCTCGAGAGATCCCGATGCCCTACAGAGCCAAATGCTTAATAGGTGACC 300  
479 TACACATGTACTTCACTCAGTTTGTGATCCGTAATATGCAATTCGAGTACTT 538  
301 TACACATGTACTTCACTCAGTTTGTGATCCGTAATATGCAATTCGAGTACTT 360  
539 CACAGTGTGTGAGAGATTAAATGAATGCTTTGTAAGCTTTTTCAGAGAGAG 598  
361 CACAGTGTGTGAGAGATTAAATGAATGCTTTGTAAGCTTTTTCAGAGAGAG 420  
599 CTTGGAGAGAGGCTTGGCGCGGCAAGACACCTCTGTGACAGAGGACCAAGGACG 658  
421 CTTGGAGAGAGGCTTGGCGCGGCAAGACACCTCTGTGACAGAGGACCAAGGACG 480  
659 ATGAAGACCCCGTGGAGCTGGCGCTCAGTGGAGTCAAGACCTCGGCTTCAGACCG 718  
481 ATGAAGACCCCGTGGAGCTGGCGCTCAGTGGAGTCAAGACCTCGGCTTCAGACCG 540  
719 TGCCGAGTGTGCTAACCGGCTCAAGGCGCAGAGCTCATATGTGATGATGACTGTTGGC 778  
541 TGCCGAGTGTGCTAACCGGCTCAAGGCGCAGAGCTCATATGTGATGATGACTGTTGGC 600  
779 AGCCCGAGAGGACCCCGGCTTACCCACCGGATTTGATTCGGGCTGGGTGAGAGGCT 838  
601 AGCCCGAGAGGACCCCGGCTTACCCACCGGATTTGATTCGGGCTGGGTGAGAGGCT 660  
839 AACAGAACGAGAGGCTGGGCAAGAGGATCAAGAGGCTTGGGGCAAGAGGAGCTGT 898  
661 AACAGAACGAGAGGCTGGGCAAGAGGATCAAGAGGCTTGGGGCAAGAGGAGCTGT 720  
899 GAGACCAACCCCTCAAGAGGCGACACCCCTCAACAGGCTTGGGGCAAGAGGAGCTGT 958  
721 GAGACCAACCCCTCAAGAGGCGACACCCCTCAACAGGCTTGGGGCAAGAGGAGCTGT 780  
959 AGACCATCAAGCAGACCCCGCTTACTGATGATGCTGTGGCTTCGGATCGAA 1018  
781 AGACCATCAAGCAGACCCCGCTTACTGATGATGCTGTGGCTTCGGATCGAA 840  
1019 GGGCGCAGCTTCGGGGCCCGCGAGATGCGAGAGGGGATGCGCAAGAGCTCGTGTCTC 1078  
841 GGGCGCAGCTTCGGGGCCCGCGAGATGCGAGAGGGGATGCGCAAGAGCTCGTGTCTC 900  
1079 TTGTGACGCGCACCTTACCTCCAGGGTGAAGCACTCGCCCGCCAGAGGAGCA 1138  
901 TTGTGACGCGCACCTTACCTCCAGGGTGAAGCACTCGCCCGCCAGAGGAGCA 960  
1139 CTGCGAGCATTTACCAAGCTGTCTCCAAAGAGCGGGGCGAGCGGAGCACTGC 1198  
961 CTGCGAGCATTTACCAAGCTGTCTCCAAAGAGCGGGGCGAGCGGAGCACTGC 1020  
1199 CAGAGTTATCTATGGGTGGGTTACACTCTTACGCGCCCTGAGAGGGGACTTCCAT 1258  
1021 CAGAGTTATCTATGGGTGGGTTACACTCTTACGCGCCCTGAGAGGGGACTTCCAT 1080  
1259 TCCCTCAACCACTGAAATGTCCTCAGACACTGTATCCAGCCACAGTGCCTCCACACA 1318  
1081 TCCCTCAACCACTGAAATGTCCTCAGACACTGTATCCAGCCACAGTGCCTCCACACA 1140  
1319 AATGGGCTTCAAGATCTCAGGCTTCCAGCTCAGAGGGTGAAGCTTCCGAGGCGCTGTG 1378  
1141 AATGGGCTTCAAGATCTCAGGCTTCCAGCTCAGAGGGTGAAGCTTCCGAGGCGCTGTG 1200  
1379 ACTTTCAGGAGCTGCTCAGTTAGCATTTACATTCACCCACGAGAGGTGGGGCC 1438  
1201 ACTTTCAGGAGCTGCTCAGTTAGCATTTACATTCACCCACGAGAGGTGGGGCC 1260  
1439 ACCCAGAAACCAAGCCCTTGGAAATGATCTTTTCAACAGGTTGCTATGGGGCC 1498  
1261 ACCCAGAAACCAAGCCCTTGGAAATGATCTTTTCAACAGGTTGCTATGGGGCC 1320



Oy	1499	ACGGGAGAGGATATAGCCCCCTTGCCAGGGTAGAGGAGAAATTCAATCAACCCAGGAAACCCCA	1558
Dd	1321	ACGGGAGAGGATATAGCCCCCTTGGCCAGGGTAGAGGAGAAATTCAATCAACCCAGGAAACCCCA	1380
Oy	1559	GGATATTAAAGAACCCCTGCTGGGGGCAACAGACATATACAGGGGGGAGTGCCTCCCT	1618
Dd	1381	GGATATTAAAGAACCCCTGCTGGGGGCAACAGACATATACAGGGGGGAGTGCCTCCCT	1440
Oy	1619	TTATTCGACACATCTCTAGTCGATTCCTGCTTTTCCCGATTGCGATTTGGGGGCC	1678
Dd	1441	TTATTCGACACATCTCTAGTCGATTCCTGCTTTTCCCGATTGCGATTTGGGGGCC	1500
Oy	1679	ACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAATTAATAGTCCCAACCCA	1738
Dd	1501	ACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAATTAATAGTCCCAACCCA	1560
Oy	1739	GGGGCGTGGGCACTTCCCAATCATTCATTTGCTTCTGCTGCCAAGTGCGAATTAAGGGGTG	1798
Dd	1561	GGGGCGTGGGCACTTCCCAATCATTCATTTGCTTCTGCTGCCAAGTGCGAATTAAGGGGTG	1620
Oy	1799	ATTGCCAACCTGG 1811	
Dd	1621	ATTGCCAACCTGG 1633	

RESULT 4	1775 bp	DNA	linear	PAT 17-JAN-2003
LOCUS	BD158346			
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.			
ACCESSION	BD158346			
VERSION	BD158346.1	GI:27864104		
KEYWORDS	UP 2002191363-A/13189.			
SOURCE	Homo sapiens (human)			

REFERENCE	(bases 1 to 175)
AUTHORS	Ota, T., Inagaki, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, T., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K., and Otsuki, T.
TITLE	Primer for synthesizing full-length cDNA and use thereof
JOURNAL	Patent: JP 2002191363-A 1189 09-JUL-2002;
COMMENT	HELIX RESEARCH INSTITUTE OS Homo sapiens (human)

PC	Location/Qualifiers	Key
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,		
PC	C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00	CC
Primer for synthesizing full-length cDNA and use thereof		FT
FT	CDS (469) .. (1275).	CDS

```

FEATURES
SOURCE
    Location/Qualifiers
    1..1775
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"

```

Query Match	89.5%	Score 1621	DB 6	Length 1775
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1621	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	191	GTGCAAGTGTGGGTTCTTGAGATCGCGGCGGTTCACACGTAGCCTTGCC	250	
Db	1	GTGCAAGTGTGGGTTCTTGAGATCGCGGCGGTTCACACGTAGCCTTGCC	60	

OY	251	AGCTCTGAGGGTGA	GTCCGTCCGGCGCGAGGTGCCCGGGACAGGCTTAA	GGCTTAA	GGCTCGCGGGGGT	310																		
Db	61	GGCTCTTGAGTGA	GTCCGTCCGGCGCGAGGTGCCCGGGACAGGCTTAA	GGCTTAA	GGCTCGCGGGGGT	120																		
OY	311	CCGGGGCCCCAGGG	ATTCCGGGCTGCAGATTGA	CGGGGATCCCGGATGCA	CCGGCGGCC	370																		
Db	121	CCGGGGCCCCAGGG	ATTCCGGGCTGCAGATTGA	CGGGGATCCCGGATGCA	CCGGCGGCC	180																		
OY	371	CCGGGCCCTCA	CCGAGCGGTCCAGA	CTGTGTGGAA	AGAGSTCGGGGACGGGTCCCTGA	430																		
Db	181	CCGGGCCCTCA	CCGAGCGGTCCAGA	CTGTGTGGAA	AGAGSTCGGGGACGGGTCCCTGA	240																		
OY	431	GGATTCGGA	TGCGCTCA	CGAGCCGA	ATGCTCA	CTTTA	GTAGTGTGAC	CTTACATATGAC	490															
Db	241	GGATTCGGA	TGCGCTCA	CGAGCCGA	ATGCTCA	CTTTA	GTAGTGTGAC	CTTACATATGAC	300															
OY	491	TTCA	CTCA	GT	TTGTGTGATCCGT	AAATGTGA	CA	AAAT	TGAAGCTACTT	CA	CAGTCTGT	550												
Db	301	TTCA	CTCA	GT	TTGTGTGATCCGT	AAATGTGA	CA	AAAT	TGAAGCTACTT	CA	CAGTCTGT	360												
OY	551	GAGA	GGATTAA	TAA	TAA	CAATGCTTT	GT	TAA	AGCTCTTT	TG	CA	AGAGGGAGGCTCGGA	CGAG	610										
Db	361	GAGA	GGATTAA	TAA	TAA	CAATGCTTT	GT	TAA	AGCTCTTT	TG	CA	AGAGGGAGGCTCGGA	CGAG	420										
OY	611	GGCGTGC	CCCGCA	AGCA	CAC	CTGTGCT	CA	C	ACAGGA	ACA	CA	GGC	CACAT	TAA	AC	CCCC	670							
Db	421	GGCGTGC	CCCGCA	AGCA	CAC	CTGTGCT	CA	C	ACAGGA	ACA	CA	GGC	CACAT	TAA	AC	CCCC	480							
OY	671	GTGAG	AGCTGAG	CCGTCA	GTGGGA	TGCAGA	CCCTCGG	CTT	CA	AGCA	CCGCTGCG	CGAG	GTGGC	730										
Db	481	GTGAG	AGCTGAG	CCGTCA	GTGGGA	TGCAGA	CCCTCGG	CTT	CA	AGCA	CCGCTGCG	CGAG	GTGGC	540										
OY	731	TAC	CGGGTCA	AGG	CA	GAG	CGTCA	TATGTGA	TGAG	CTCT	TT	TG	CA	AG	CC	CA	CGC	790						
Db	541	TAC	CGGGTCA	AGG	CA	GAG	CGTCA	TATGTGA	TGAG	CTCT	TT	TG	CA	AG	CC	CA	CGC	600						
OY	791	A	CCCGG	CT	CA	CCCA	CGG	ACTT	GCAT	CCG	CT	CG	GT	GTGA	GA	AG	GG	CTA	CA	GA	CA	CA	GA	850
Db	601	A	CCCGG	CT	CA	CCCA	CGG	ACTT	GCAT	CCG	CT	CG	GT	GTGA	GA	AG	GG	CTA	CA	GA	CA	CA	GA	660
OY	851	GGCGTGG	CA	AG	AG	GATCGA	AGG	CCCTT	GG	GGG	CA	AA	GG	GG	AG	CT	GTGA	ACA	CCCC	910				
Db	661	GGCGTGG	CA	AG	AG	GATCGA	AGG	CCCTT	GG	GGG	CA	AA	GG	GG	AG	CT	GTGA	ACA	CCCC	720				
OY	911	TCA	AGGGG	CA	GA	CA	CCCCCA	CCCTCA	CA	CA	AG	AA	GA	AA	CA	AA	TACAG	CCCAT	AGC	970				
Db	721	TCA	AGGGG	CA	GA	CA	CCCCCA	CCCTCA	CA	CA	AG	AA	GA	AA	CA	AA	TACAG	CCCAT	AGC	780				
OY	971	CAC	AC	CCCG	CTTACTGT	GATGAGT	CGCTG	TT	GT	TC	CCGAT	CTGA	AG	GG	CG	CA	CTTC	1030						
Db	781	CAC	AC	CCCG	CTTACTGT	GATGAGT	CGCTG	TT	GT	TC	CCGAT	CTGA	AG	GG	CG	CA	CTTC	840						
OY	1031	GGGG	CCCCCG	GA	ATGG	GA	AGGG	GA	TCG	CGA	AGTCC	CGT	CT	CTT	TG	GA	GGCA	1090						
Db	841	GGGG	CCCCCG	GA	ATGG	GA	AGGG	GA	TCG	CGA	AGTCC	CGT	CT	CTT	TG	GA	GGCA	900						
OY	1091	CCA	CTTAC	CCCC	CA	AGGG	GTAG	CA	CTCG	CCCG	CCCC	CA	AGG	AG	CA	CA	CTCG	CGA	CTT	1150				
Db	901	CCA	CTTAC	CCCC	CA	AGGG	GTAG	CA	CTCG	CCCG	CCCC	CA	AGG	AG	CA	CA	CTCG	CGA	CTT	960				
OY	1151	CAC	CC	AG	CTG	ATC	CTTCA	AG	CA	AG	AG	CG	GGG	CA	AG	CG	CA	CTC	CA	GA	AG	TATCT	1210	
Db	961	CAC	CC	AG	CTG	ATC	CTTCA	AG	CA	AG	AG	CG	GGG	CA	AG	CG	CA	CTC	CA	GA	AG	TATCT	1020	
OY	1211	ATGG	GTGG	GT	TAC	CTCTT	CA	CG	CCCC	CT	GAA	G	CGGG	GA	CTTT	CC	CA	TTT	CC	CT	CA	CC	CA	1270
Db	1021																							

Db 1141 GATCTCAGGCGCTTCCAGTCAGGGGTGACCTTCCGAGGCCCGCTGAGCTTCCAGGGCT 1200  
Qy 1391 CGCTCAAGTTAGCATTTAGTGCATCTTACCCCAAGAGTGGGGGCAACCGAAGAACCA 1450  
Db 1201 CGCTCAAGTTAGCATTTAGTGCATCTTACCCCAAGAGTGGGGGCAACCGAAGAACCA 1260  
Qy 1451 AAGCCCTTGAAGATGATCTTTTCATCAGGGTGTCTTATGGGGGCAAGGCGACAGT 1510  
Db 1261 AAGCCCTTGAAGATGATCTTTTCATCAGGGTGTCTTATGGGGGCAAGGCGACAGT 1320  
Qy 1511 ATGGCCCTTGGCAGGGTAGAGAGACATTCATCCAGGGAACCCAGGATTTAAAGAA 1570  
Db 1321 ATGGCCCTTGGCAGGGTAGAGAGACATTCATCCAGGGAACCCAGGATTTAAAGAA 1380  
Qy 1571 GCGCCCTTGGGGGAGACACACATAGAGAGGGTGGGAGTCCCTCTTATCTTGACAA 1630  
Db 1381 GCGCCCTTGGGGGAGACACACATAGAGAGGGTGGGAGTCCCTCTTATCTTGACAA 1440  
Qy 1631 TCTCTAGTCGATCTTCTCTCTTCTCTCCGATTCGAGATTGGGGGCACTCTAAGATG 1690  
Db 1441 TCTCTAGTCGATCTTCTCTCTTCTCTCCGATTCGAGATTGGGGGCACTCTAAGATG 1500  
Qy 1691 CCTCTCTCAGCCCTGTCTCAACCATTAATGTCGCAACCCAGGGGCTGTGAC 1750  
Db 1501 CCTCTCTCAGCCCTGTCTCAACCATTAATGTCGCAACCCAGGGGCTGTGAC 1560  
Qy 1751 CTCTCCATCATCATCTGTCTGTGTGCAAGTGGAAATTAACGGCGGATTCGCAACCTG 1810  
Db 1561 CTCTCCATCATCATCTGTGTGTGCAAGTGGAAATTAACGGCGGATTCGCAACCTG 1620  
Qy 1811 G 1811  
Db 1621 G 1621

RESULT 5  
AX880378 1775 bp DNA linear PAT 17-DEC-2003  
LOCUS AX880378  
DEFINITION Sequence 15283 from Patent EP1074617.  
ACCESSION AX880378  
VERSION AX880378.1 GI:40035114  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukayote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Makamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primers for synthesizing full-length cDNA and their use  
JOURNAL Patent: EP 1074617-A 15283 07-FEB-2001  
Research Association for Biotechnology (JRP)  
FEATURES  
source  
1. 1775  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
467..1278  
/note="unassigned protein product"  
/codon\_start=1  
/protein\_id="CAB90924.1"  
/db\_xref="GI:40035115"  
/translation="MKTPELVASGMOTLGLHRCRGYRVKARTSYVDETLLFGSPAG  
TRTPPDPDPVWEKANKRTGVGKASKALDAKSCSTTTSRGSFTPLTRKKNKRP  
ISHTPSYCDLSLFGSRBEGAFAPRMKADALRLALWTPPTPGSHSPRRKAP  
LRAIHAPGSKTEGPAADSQKLSMGSLHSRPLKRLSHSLTLANPSTGHATSAP  
HTNGPQLRSTSGVTPRSPPLVTSRARSVISVSTPRRGATQKPKPWK"

ORIGIN  
Query Match 89.5%; Score 1621; DB 6; Length 1775;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 GTGCAAGTGTGGGTTCTGGTTCTGTGATTCGCGGCGCTTCAACGTAAGCTGTGCC 250  
Db 1 GTGCAAGTGTGGGTTCTGGTTCTGTGATTCGCGGCGCTTCAACGTAAGCTGTGCC 60  
Qy 251 GGTCTCTCGGGTGAATTCGGTCCCGCGCGGTGCTCCCGGACGGCTTAAAGTGTCCGGGGGT 310  
Db 61 GGTCTCTCGGGTGAATTCGGTCCCGCGCGGTGCTCCCGGACGGCTTAAAGTGTCCGGGGGT 120  
Qy 311 CGGGGGCCCGAGGATTCGGGGGTGAGATGACGGGGATCCCGGATGACCGCGGCGCC 370  
Db 121 CGGGGGCCCGAGGATTCGGGGGTGAGATGACGGGGATCCCGGATGACCGCGGCGCC 180  
Qy 371 CCGGCGCTTCAACGAGGGTCCAGACTGTGGGAAGAGGTGCGGGGACGGGTCTCTGA 430  
Db 181 CCGGCGCTTCAACGAGGGTCCAGACTGTGGGAAGAGGTGCGGGGACGGGTCTCTGA 240  
Qy 431 GATCCCGATGCTTCAAGCCAAAGATGCTCACTTTATAGTGTGACTTACATGTGAC 490  
Db 241 GATCCCGATGCTTCAAGCCAAAGATGCTCACTTTATAGTGTGACTTACATGTGAC 300  
Qy 491 TTGACCTGATTTGTGATCCGTAATAGGCAATTCGAAGCTACTTCAAGTGTGCTT 550  
Db 301 TTGACCTGATTTGTGATCCGTAATAGGCAATTCGAAGCTACTTCAAGTGTGCTT 360  
Qy 551 GAGAGATTAAATGAACAATGCTGTAAAGCTCTTTGACAGAGGAGGCTCGGAAGCAG 610  
Db 361 GAGAGATTAAATGAACAATGCTGTAAAGCTCTTTGACAGAGGAGGCTCGGAAGCAG 420  
Qy 611 GGCCTGGCGGACAGACACCTGTGTGACAGAGGACACAGGCGACATGAAGACCCCG 670  
Db 421 GGCCTGGCGGACAGACACCTGTGTGACAGAGGACACAGGCGACATGAAGACCCCG 480  
Qy 671 GTGAGCTGGCGCTGAGTGGGATGACAGCCCTGGCGCTTCAAGCAGGCTCGAGGTGGC 730  
Db 481 GTGAGCTGGCGCTGAGTGGGATGACAGCCCTGGCGCTTCAAGCAGGCTCGAGGTGGC 540  
Qy 731 TACCGGATCAAGGCGCAGACGTCATATGAGATGAGACTCTGTTGGACGCGCAGACAGC 790  
Db 541 TACCGGATCAAGGCGCAGACGTCATATGAGATGAGACTCTGTTGGACGCGCAGACAGC 600  
Qy 791 ACCCGGCTTACCCACCGGACTTGTGATTCGGCGCTTGGGTGAGAGGCTTAAAGACACAGA 850  
Db 601 ACCCGGCTTACCCACCGGACTTGTGATTCGGCGCTTGGGTGAGAGGCTTAAAGACACAGA 660  
Qy 851 GGGGTGGGCAAGAGGATGAGAGGCTTGGGGGCAAGGGGAGCTGTGAGACACACCCG 910  
Db 661 GGGGTGGGCAAGAGGATGAGAGGCTTGGGGGCAAGGGGAGCTGTGAGACACACCCG 720  
Qy 911 TCAAGGGGCGACACCCGACCTTCAACCAAGAGAGAGAAATTAAGACCCATCAGC 970  
Db 721 TCAAGGGGCGACACCCGACCTTCAACCAAGAGAGAGAAATTAAGACCCATCAGC 780  
Qy 971 CACACCCGCTTACTGTGTGATGATGCTGTGTGCTCCGACTGTGAGGCGCAGCTTC 1030  
Db 781 CACACCCGCTTACTGTGTGATGATGCTGTGTGCTCCGACTGTGAGGCGCAGCTTC 840  
Qy 1031 GGGGGCCCGGAGTGGCGAAGGGGGATGCGGCAAGCTCGTGTGAGAGGCCA 1090  
Db 841 GGGGGCCCGGAGTGGCGAAGGGGGATGCGGCAAGCTCGTGTGAGAGGCCA 900  
Qy 1091 CCACTTACCCCGAGGGGTAGCCTGCGCCCGGCGGAGGAGGACCACTGCGAGGCATT 1150  
Db 901 CCACTTACCCCGAGGGGTAGCCTGCGCCCGGCGGAGGAGGACCACTGCGAGGCATT 960  
Qy 1151 CACCCAGCTGTCTTCCCAAGACAGAGCCGGGCGCAGCGGCAAGCTCCGAGAGTTATCT 1210  
Db 961 CACCCAGCTGTCTTCCCAAGACAGAGCCGGGCGCAGCGGCAAGCTCCGAGAGTTATCT 1020  
Qy 1211 ATGGGTGGGTAACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 1270  
Db 1021 ATGGGTGGGTAACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 1080

```

Oy 1271 CTGAATGTCCTCCAGCAGCTGCTATCCAGCCAGTGCCTCCACAAATGGCTCTG 1330
Db 1081 CTGAATGTCCTCCAGCAGCTGCTATCCAGCCAGTGCCTCCACAAATGGCTCTG 1140
Oy 1331 GATCTCAGGCGCTTCCAGCTCAGGGGAGCTTCCGAGCGCTCTGCTGCTTCCAGGGCT 1390
Db 1141 GATCTCAGGCGCTTCCAGCTCAGGGGAGCTTCCGAGCGCTCTGCTGCTTCCAGGGCT 1200
Oy 1391 CGCTCAGTTAGCATTTTCAGTGCATCTACCCACGACGAGTGGGGCCACCCAGAAACCA 1450
Db 1201 CGCTCAGTTAGCATTTTCAGTGCATCTACCCACGACGAGTGGGGCCACCCAGAAACCA 1260
Oy 1451 AAGCCCTTGGAAATGATATCTTTTCATCAGGGTTCCTATGCGGCCACGCGCAGAGT 1510
Db 1261 AAGCCCTTGGAAATGATATCTTTTCATCAGGGTTCCTATGCGGCCACGCGCAGAGT 1320
Oy 1511 ATGGCCCTTGGCAGGATGAGGAGCATTCATCACTCCAGGAAACCCAGAGTATTAAAGAA 1570
Db 1321 ATGGCCCTTGGCAGGATGAGGAGCATTCATCACTCCAGGAAACCCAGAGTATTAAAGAA 1380
Oy 1571 GCCCTGTGGGGGAGCAGACATAGCAGGGGTGGGAGTGCCTCCCTTATCTGACAA 1630
Db 1381 GCCCTGTGGGGGAGCAGACATAGCAGGGGTGGGAGTGCCTCCCTTATCTGACAA 1440
Oy 1631 TCTCTAGTGCATTTCTTCCCTTTTCTCCGATTTGCGGATTTGGGGGCCACTCTTAAGATG 1690
Db 1441 TCTCTAGTGCATTTCTTCCCTTTTCTCCGATTTGCGGATTTGGGGGCCACTCTTAAGATG 1500
Oy 1691 CCTCTCTCAGGCGCTCTCTCAACCATCTCCAAATTAAGTCCACCCAGGGGCTTGGCAC 1750
Db 1501 CCTCTCTCAGGCGCTCTCTCAACCATCTCCAAATTAAGTCCACCCAGGGGCTTGGCAC 1560
Oy 1751 CTCCTCAGCATTCATCTGCTGCTGCTGCGCAAGTGGCAATTAACGGGCTGATGGCAACCTG 1810
Db 1561 CTCCTCAGCATTCATCTGCTGCTGCTGCGCAAGTGGCAATTAACGGGCTGATGGCAACCTG 1620
Oy 1811 G 1811
Db 1621 G 1621

RESULT 6
AK027733 1775 bp mRNA linear PRI 30-JAN-2004
LOCUS Homo sapiens cDNA FLJ14827 fls, clone OVARC100086.
DEFINITION AK027733
ACCESSION AK027733.1 GI:14042629
VERSION AK027733.1
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakematsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahashi, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimura, M., Watanabe, M.,
Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hottel, T., Kusano, J., Kanehori, K., Takahashi, F., Hara, R.,
Tateuchi, K., Arita, M., Imose, N., Musashino, K., Yunki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Satoh, N., Takami, S.,
Shiohata, N., Sano, S., Moriya, S., Motiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,

```

```

Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hiro, M., Omori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikem, Y., Okamoto, S., Okitani, R., Kawakami, K.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T.,
Oyama, M., Hara, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Maubou, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
JOURNAL
PUBMED
14702039
REFERENCE
AUTHORS
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakami, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakematsu, A.,
Nakamura, Y., Nagahashi, K., Maubou, Y., Ninomiya, K. and Iwayanagi, T.
NEO human cDNA sequencing project
Unpublished
3 (bases 1 to 1775)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction,
5'- & 3'-end and one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
location/Qualifiers
1..1775
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OVARC100086"
/tissue_type="ovary, tumor tissue"
/clone_lib="OVARC1"
/note="Cloning vector: pME18SF13"
469..1278
/note="unamed protein product"
/codon_start=1
/db_xref="GI:14042630"
/translation="MKTPVELAVSGMOTLGLQHCRCGRYRKARTSYDFTLFGSPAG
TRTPPDPPEVWVKANRTRGVGEASALGAKGSCETTPSRGSPITLTPPKKKYRP
ISHTPSYCDSELPFSRSEGSFAPRNAGKDAALRLMTPTPTPSHSRPREAP
LRAIHPGSPKTEPSPADSOGLSMGSLHSRPLKRGSHLTLILNVSTGHPATSAP
HTNPDPDRPETSQVTRPSPLVTRASVSISVSTPRRGATQKPKPKPK"
ORIGIN
Query Match 89.5%; Score 1621; DB 9; Length 1775;
Best local similarity 100.0%; Pred. No. 0;
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 191 GTGCAAAAGTGTGGGTTTGGGTTTTCGATTTCCGGGCGCTTACAGTACCTGTGCC 250
Db 1 GTGCAAAAGTGTGGGTTTGGGTTTTCGATTTCCGGGCGCTTACAGTACCTGTGCC 60
Oy 251 GGCTCTCTGGGTGAGTCCGTCCGCGCGCGCGCGCGCGCGCTAGCTGCGGGGGT 310
Db 61 GGCTCTCTGGGTGAGTCCGTCCGCGCGCGCGCGCGCGCGCTAGCTGCGGGGGT 120
Oy 311 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 370
Db 121 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180

```

QY 371 CCGGCGCCCTGACCGAGCGGTCACAGCTGAGGAGAGAGGTCGAGGAGCGGTCCTGA 430  
 Db 181 CCGGCGCCCTGACCGAGCGGTCACAGCTGAGGAGAGAGGTCGAGGAGCGGTCCTGA 240  
 QY 431 GGATCCCGATGCTCAGAGCGGAGATGCTGAGCTTTAAGGTGACCTTACATGTGAC 490  
 Db 241 GGATCCCGATGCTCAGAGCGGAGATGCTGAGCTTTAAGGTGACCTTACATGTGAC 300  
 QY 491 TTGACCTCAGTTTGTATCCGTAAATGACAAATTCGAAAGCTACTTCCAGGCTGT 550  
 Db 301 TTGACCTCAGTTTGTATCCGTAAATGACAAATTCGAAAGCTACTTCCAGGCTGT 360  
 QY 551 GAGGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGACAGAGAGAGCTTCGGAACAG 610  
 Db 361 GAGGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGACAGAGAGAGCTTCGGAACAG 420  
 QY 611 GGCCTGCGCGGACAGACACCTGCTGTACACAGAGAACACAGGACGATGAACACCCC 670  
 Db 421 GGCCTGCGCGGACAGACACCTGCTGTACACAGAGAACACAGGACGATGAACACCCC 480  
 QY 671 GTGAGAGCTGCGCTCAGTGGAGATGACACCTCGGCTTACAGACACGCTGTCCAGGTGAC 730  
 Db 481 GTGAGAGCTGCGCTCAGTGGAGATGACACCTCGGCTTACAGACACGCTGTCCAGGTGAC 540  
 QY 731 TACCGGATCAAGGCGCAGAGCGTCATATGTGATGAGACTGTTTGGACAGCCAGAGG 790  
 Db 541 TACCGGATCAAGGCGCAGAGCGTCATATGTGATGAGACTGTTTGGACAGCCAGAGG 600  
 QY 791 ACCGCGCTTACCCAGCGGATTCGATCCGCTTGGGTGAGAGAGGCTTACAGAACAGA 850  
 Db 601 ACCGCGCTTACCCAGCGGATTCGATCCGCTTGGGTGAGAGAGGCTTACAGAACAGA 660  
 QY 851 GCGCTGGGCAAGAGGAGATGGAAGGCTTGGGGGCAAGAGGAGAGCTGTGAGACACCCC 910  
 Db 661 GCGCTGGGCAAGAGGAGATGGAAGGCTTGGGGGCAAGAGGAGAGCTGTGAGACACCCC 720  
 QY 911 TCAAGGGGCAAGACCCCACTTCAACAAGAGAGAGAGCAATATCAGACCCATGAGC 970  
 Db 721 TCAAGGGGCAAGACCCCACTTCAACAAGAGAGAGAGCAATATCAGACCCATGAGC 780  
 QY 971 CACACCCGCTTACCTGATGAGTCCGCTGTTGGCTCCGATCTGAAGGCGCCAGCTTC 1030  
 Db 781 CACACCCGCTTACCTGATGAGTCCGCTGTTGGCTCCGATCTGAAGGCGCCAGCTTC 840  
 QY 1031 GGGGCGCGCGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 1090  
 Db 841 GGGGCGCGCGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 900  
 QY 1091 CCACTTACCCCAAGGGGATGACCACTGCGCCCGCCCAAGGAGAGGAGAGGAGAGGAG 1150  
 Db 901 CCACTTACCCCAAGGGGATGACCACTGCGCCCGCCCAAGGAGAGGAGAGGAGAGGAG 960  
 QY 1151 CACCCAGCTGCTCCCTCCAGAGACAGAGCGGGGCGAGCGGAGAGCTTCCAGAGTTATCT 1210  
 Db 961 CACCCAGCTGCTCCCTCCAGAGACAGAGCGGGGCGAGCGGAGAGCTTCCAGAGTTATCT 1020  
 QY 1211 ATGGGTGGGTACACTCTTCAAGCGCCCTGAAAGGGGGGAGCTTCCACTTCCCTACCCAC 1270  
 Db 1021 ATGGGTGGGTACACTCTTCAAGCGCCCTGAAAGGGGGGAGCTTCCACTTCCCTACCCAC 1080  
 QY 1271 CTGATGTCCCAAGCACTGATCATCAGACCAAGTGCCCGCCCAAGAAATGAGGCTCAG 1330  
 Db 1081 CTGATGTCCCAAGCACTGATCATCAGACCAAGTGCCCGCCCAAGAAATGAGGCTCAG 1140  
 QY 1331 GATTCAGGCTTCCAGCTCAGGAGGAGCTTCCGAGAGCCCTGAGTGAATTCAGAGGCT 1390  
 Db 1141 GATTCAGGCTTCCAGCTCAGGAGGAGCTTCCGAGAGCCCTGAGTGAATTCAGAGGCT 1200  
 QY 1391 CGCTCAGTTGACATTTAGTGCCATCTAACCCAGAGAGGAGGAGGAGGAGGAGGAGGAG 1450  
 Db 1201 CGCTCAGTTGACATTTAGTGCCATCTAACCCAGAGAGGAGGAGGAGGAGGAGGAGGAG 1260  
 QY 1451 AAGCCCTTGGAAATGATTAATCTTTTCAATCAGGCTTCTATGGGCGACAGGAGAGGT 1510

Db 1261 AAGCCCTTGGAAATGATTAATCTTTTCAATCAGGCTTCTATGGGCGACAGGAGAGGT 1320  
 QY 1511 ATGGCCCTTGGAGGAGGATGAGAGACATTCATATCCAGAGAACCCAGATTAAGA 1570  
 Db 1321 ATGGCCCTTGGAGGAGGATGAGAGACATTCATATCCAGAGAACCCAGATTAAGA 1380  
 QY 1571 GCGCCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1630  
 Db 1381 GCGCCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
 QY 1631 TCTCTAGTCGATTCCTTGGCTTTTCTCCGATTTGGGAGCCACTTAAAGATG 1690  
 Db 1441 TCTCTAGTCGATTCCTTGGCTTTTCTCCGATTTGGGAGCCACTTAAAGATG 1500  
 QY 1691 CCTCTCTCCAGCCCTGTCTCAACATCTCAATTAATAGTCCAAACCCAGGAGGCTGAC 1750  
 Db 1501 CCTCTCTCCAGCCCTGTCTCAACATCTCAATTAATAGTCCAAACCCAGGAGGCTGAC 1560  
 QY 1751 CTCGCAATCATCATTTGCTTGTCTGCAAGTGCAGATTAACCGGCTGATTCGCAACTG 1810  
 Db 1561 CTCGCAATCATCATTTGCTTGTCTGCAAGTGCAGATTAACCGGCTGATTCGCAACTG 1620  
 QY 1811 G 1811  
 Db 1621 G 1621  
 RESULT 7  
 AK123762  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1  
 Oca.T., Suzuki.Y., Nishikawa.T., Otsuki.T., Sugiyama.T., Irie.R.,  
 Wakamatsu.A., Hayashi.K., Sato.H., Nagai.K., Kimura.K., Makita.H.,  
 Sekine.M., Obayashi.M., Nishi.T., Shibahara.T., Tanaka.T.,  
 Ishii.S., Yamamoto.J., Saito.K., Kawai.Y., Isono.Y., Nakamura.Y.,  
 Nagahari.K., Murakami.K., Yasuda.T., Iwayanagi.T., Wagatsuma.M.,  
 Shiratori.A., Sudo.H., Hosoi.T., Kaku.Y., Kodaira.H., Kondo.H.,  
 Sugawara.M., Takahashi.M., Kanda.K., Yokoi.T., Furuya.T.,  
 Kikkawa.E., Omura.Y., Abe.K., Kamihara.K., Katsuta.N., Sato.K.,  
 Tanikawa.M., Yamazaki.M., Niinomi.Y., Ishibashi.T., Yamauchi.H.,  
 Murakawa.K., Fujimori.K., Tanai.H., Kimata.M., Watanabe.M.,  
 Hirakawa.S., Chiba.Y., Ishida.S., Ono.Y., Takiguchi.S., Watanabe.S.,  
 Yoshida.M., Houta.T., Kusano.J., Kanehori.K., Takahashi-Fujii.A.,  
 Hara.H., Tanase.T., Nomura.Y., Togiya.S., Komai.F., Hara.R.,  
 Takeuchi.K., Arita.M., Imose.N., Mutsaers.K., Yuuki.H., Oshima.A.,  
 Sasaki.N., Aotuka.S., Yoshikawa.Y., Matsunawa.H., Ichihara.T.,  
 Shiohata.N., Sano.S., Moriya.S., Momiya.A., Satoh.N., Takami.S.,  
 Terahima.Y., Suzuki.O., Nakagawa.S., Senoh.A., Mizoguchi.H.,  
 Goto.Y., Shimizu.F., Wakebe.H., Hishigaki.H., Watanabe.T.,  
 Sugiyama.Y., Takemoto.M., Kawakami.B., Yanahashi.M., Watanabe.K.,  
 Kumagai.A., Itakura.S., Fukuzumi.Y., Fujimori.Y., Komiya.M.,  
 Tashiro.H., Tanigami.A., Fujiwara.T., Ono.T., Yamada.K., Fujii.Y.,  
 Ozaki.K., Hiro.M., Ohmori.Y., Kawabata.A., Hiki.T., Kobatake.N.,  
 Inagaki.H., Ikema.Y., Okamoto.S., Okitani.R., Kawakami.K.,  
 Noguchi.S., Itoh.T., Shigeta.K., Senba.T., Matsuura.K.,  
 Nakajima.Y., Mizuno.T., Morinaga.M., Sasaki.M., Togashi.T.,  
 Oyama.M., Hata.H., Watanabe.M., Komatsu.T., Mizushima-Sugano.J.,  
 Satoh.T., Shira.Y., Takahashi.Y., Nakagawa.K., Okumura.K.,  
 Nagase.T., Nomura.N., Kikuchi.H., Masuno.Y., Yamashita.R.,  
 Nakai.K., Yada.T., Nakamura.Y., Ohara.O., Isegaki.T. and Sugano.S.  
 Complete sequencing and characterization of 21,243 full-length  
 human cDNAs  
 JOURNAL  
 PUBMED  
 Nat. Genet. 36 (1), 40-45 (2004)  
 14702039

REFERENCE  
AUTHORS  
2  
Tashiro, H., Yamazaki, M., Matanabe, K., Kumagai, A., Itakura, S.,  
Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R.,  
Otsuki, T., Sato, H., Makamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,  
Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,  
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,  
Magatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,  
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
Nagahata, K., Mashino, Y., Nagai, K., and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 1709)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan: cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
Location/Qualifiers  
1. 1709  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMR322006520"  
/cell\_line="IMR32"  
/cell\_type="neuroblastoma"  
/clone\_id="IMR32"  
/note="Cloning vector: pWE18FLJ3"

ORIGIN  
Query Match 88.2%; Score 1597.6; DB 9; Length 1709;  
Best Local Similarity 94.5%; Pred. No. 0;  
Matches 1705; Conservative 0; Mismatches 4; Indels 95; Gaps 1;

QY 1 AGTGCCTCGGCGCCCTCGGCGGCTCTAGTACACACGACCTGAGTGTGAGACCAAGAGAC 60  
DB 1 AGTGCCTCGGCGCCCTCGGCGGCTCTAGTACACACGACCTGAGTGTGAGACCAAGAGAC 60  
QY 61 CCTCTCATGTTTAAAGGACCTCTCTGAGCTTCAAGAGCGTGGCGCCGCTCTGGCGGAC 120  
DB 61 CCTCTCATGTTTAAAGGACCTCTCTGAGCTTCAAGAGCGTGGCGCCGCTCTGGCGGAC 120  
QY 121 TCCGCCCATTCGCGGGGCGGAATGGTCCGCGTCCGCGAGTCTGCTGCTCTCC 180  
DB 121 TCCGCCCATTCGCGGGGCGGAATGGTCCGCGTCCGCGAGTCTGCTGCTCTCC 180  
QY 121 TCCGCCCATTCGCGGGGCGGAATGGTCCGCGTCCGCGAGTCTGCTGCTCTCC 180  
DB 121 TCCGCCCATTCGCGGGGCGGAATGGTCCGCGTCCGCGAGTCTGCTGCTCTCC 180  
QY 181 TGGTTGCGGGGTGCAAGTGGTCTGCGGTTCTGGAATTCGCGGGCCGTTCAACAGT 240  
DB 181 TGGTTGCGGGGTGCAAGTGGTCTGCGGTTCTGGAATTCGCGGGCCGTTCAACAGT 240  
QY 181 TGGTTGCGGGGTGCAAGTGGTCTGCGGTTCTGGAATTCGCGGGCCGTTCAACAGT 240  
DB 181 TGGTTGCGGGGTGCAAGTGGTCTGCGGTTCTGGAATTCGCGGGCCGTTCAACAGT 240  
QY 241 AGCCGTGTCGCGCTCTCTCGGAGTCCGTCGCGCGGAGTCCCGGAGCGACTTAAGC 300  
DB 241 AGCCGTGTCGCGCTCTCTCGGAGTCCGTCGCGCGGAGTCCCGGAGCGACTTAAGC 300  
QY 241 AGCCGTGTCGCGCTCTCTCGGAGTCCGTCGCGCGGAGTCCCGGAGCGACTTAAGC 300  
DB 241 AGCCGTGTCGCGCTCTCTCGGAGTCCGTCGCGCGGAGTCCCGGAGCGACTTAAGC 300  
QY 301 TGCAGGGGGTCCGGGGGCCCCAGGACATTCGCGGCTGCAGATTGACGCGGATCCCGATGCA 360  
DB 301 TGCAGGGGGTCCGGGGGCCCCAGGACATTCGCGGCTGCAGATTGACGCGGATCCCGATGCA 360  
QY 258 -----CGGATGCA 265  
DB 258 -----CGGATGCA 265  
QY 361 CCGCGCGCCCGCGGCGCTCAACGAGGATCCAGACTGTTGGAGGAAGAGTGGGGAGC 420  
DB 361 CCGCGCGCCCGCGGCGCTCAACGAGGATCCAGACTGTTGGAGGAAGAGTGGGGAGC 420  
QY 266 CCGCGCGCCCGCGGCGCTCAACGAGGATCCAGACTGTTGGAGGAAGAGTGGGGAGC 325  
DB 266 CCGCGCGCCCGCGGCGCTCAACGAGGATCCAGACTGTTGGAGGAAGAGTGGGGAGC 325  
QY 421 GGGTTCCTGAGAGATCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGAGCTTA 480  
DB 421 GGGTTCCTGAGAGATCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGAGCTTA 480  
QY 326 GGGTTCCTGAGAGATCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGAGCTTA 385  
DB 326 GGGTTCCTGAGAGATCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGAGCTTA 385

QY 481 CAGATGATCTTACCTCAGTTTGTGATCCGTAATAATGACCAATTCGAGCTACTTCA 540  
DB 386 CAGATGATCTTACCTCAGTTTGTGATCCGTAATAATGACCAATTCGAGCTACTTCA 445  
QY 541 CAGTCTGTGAGAGATTAATAATGAAACAATGCTTTGTAAGCTCTTTGCGAGGAGGCC 600  
DB 446 CAGTCTGTGAGAGATTAATAATGAAACAATGCTTTGTAAGCTCTTTGCGAGGAGGCC 505  
QY 601 TCGAAGACGAGGCGCTGCGCGGACGACACCTGCTGTCAACGAGGACCAAGGACAT 660  
DB 506 TCGAAGACGAGGCGCTGCGCGGACGACACCTGCTGTCAACGAGGACCAAGGACAT 565  
QY 661 GAAAGCCCCCGTGGAGTGGCGGTGAGTGGAGTGGAGCCCTCGGCTTTCAGCACCGCTG 720  
DB 566 GAAAGCCCCCGTGGAGTGGCGGTGAGTGGAGTGGAGCCCTCGGCTTTCAGCACCGCTG 625  
QY 721 CCGAGTGTGCTTACCGGCTCAAGGCGCAGGACGTCATATGTGATGAGTCTGTTGGAG 780  
DB 626 CCGAGTGTGCTTACCGGCTCAAGGCGCAGGACGTCATATGTGATGAGTCTGTTGGAG 685  
QY 781 CCGAGGAGCACCGGCTTACCCACCGGACTTGATCCGCTGCGGTGAGAGAGCTTAA 840  
DB 686 CCGAGGAGCACCGGCTTACCCACCGGACTTGATCCGCTGCGGTGAGAGAGCTTAA 745  
QY 841 CAGAACGAGGCGTGGGCAAGGAGGAGCATGCAAGGCTTGGGGGCAAGGGAGCTTGA 900  
DB 746 CAGAACGAGGCGTGGGCAAGGAGGAGCATGCAAGGCTTGGGGGCAAGGGAGCTTGA 805  
QY 901 GACCAACCCCTCAAGGGGAGAGACACCCCACTTCAACCAAGAGAGAGCAATTAAG 960  
DB 806 GACCAACCCCTCAAGGGGAGAGACACCCCACTTCAACCAAGAGAGAGCAATTAAG 865  
QY 961 ACCCATGACCAACCCCGTCTTACTGATGATGCTGCTGTTGGCTCCCAATCGAAG 1020  
DB 866 ACCCATGACCAACCCCGTCTTACTGATGATGCTGCTGTTGGCTCCCAATCGAAG 925  
QY 1021 CGCAGCTTGGGGGCGCGGAGATGGGAGGAGATGCCCAAGCTCCGCTCTCTT 1080  
DB 926 CGCAGCTTGGGGGCGCGGAGATGGGAGGAGATGCCCAAGCTCCGCTCTCTT 985  
QY 1081 GTGAGCGCACACACTTACCCCAAGGGGAGGACCTCGCGCGCCAGGAGGAGCACACT 1140  
DB 986 GTGAGCGCACACACTTACCCCAAGGGGAGGACCTCGCGCGCCAGGAGGAGCACACT 1045  
QY 1141 GCGAGCAATTCACCGAGTGGTCCCTCCAGACAGAGCGGGGCCAGCGGAGACTCCA 1200  
DB 1046 GCGAGCAATTCACCGAGTGGTCCCTCCAGACAGAGCGGGGCCAGCGGAGACTCCA 1105  
QY 1201 GAAGTTATCTATGGGTGGGTTTACACTCTTCAACGCGCCCTGAAGGCGGAGACTTCCATTTC 1260  
DB 1106 GAAGTTATCTATGGGTGGGTTTACACTCTTCAACGCGCCCTGAAGGCGGAGACTTCCATTTC 1165  
QY 1261 CTTACACCACTGATATGCCCCAGACATGATCAGACCAAGTGGCCCCCAACAAA 1320  
DB 1166 CTTACACCACTGATATGCCCCAGACATGATCAGACCAAGTGGCCCCCAACAAA 1225  
QY 1321 TGGGCTTCAAGATCTCAGGCTTCCACGTCAGGGGTGACCTTCCGAGCCCTGCTGTAC 1380  
DB 1226 TGGGCTTCAAGATCTCAGGCTTCCACGTCAGGGGTGACCTTCCGAGCCCTGCTGTAC 1285  
QY 1381 TTCCAGGGCTGCTCAGTATGATTTTCAATGATGTCATACCCCAAGAGAGTGGGGCCAC 1440  
DB 1286 TTCCAGGGCTGCTCAGTATGATTTTCAATGATGTCATACCCCAAGAGAGTGGGGCCAC 1345  
QY 1441 CCAAGAACCAAGCCCCCTTGAATGATCTTCTTCAAGGGGTGCTATGAGGGGCCAC 1500  
DB 1346 CCAAGAACCAAGCCCCCTTGAATGATCTTCTTCAAGGGGTGCTATGAGGGGCCAC 1405  
QY 1501 GCGACAGATATGCGCCCTTCCAGGAGTGAAGACATTCATCAACCAAGGAGCCCAAG 1560  
DB 1406 GCGACAGATATGCGCCCTTCCAGGAGTGAAGACATTCATCAACCAAGGAGCCCAAG 1465  
QY 1561 TATTAAAGAGCCCTGTGGGGGAGACAGACATGACAGGGGTGGGACATGCTCCTTT 1620  
DB 1561 TATTAAAGAGCCCTGTGGGGGAGACAGACATGACAGGGGTGGGACATGCTCCTTT 1620



Db	1466	TATTAAACAGCCCTGTGGGGGGGACAGACATGACAGGGGTGGGCAGTGCTCTCTTT	15225
Oy	1621	ATTCGACAATCTCTAGTCGATATTTCTGGCTTTTCTCCGATTTGGGATTTGGGGCCAC	1680
Db	1526	ATCCGACAAATCTCTAGTCGATATTTCTGGCTTTTCTCCGATTTGGGATTTGGGGCCAC	1585
Oy	1661	CTTAAAGATGCTCTCTCCAGCCCTGTGTCTCAACCAATCTCCAAATTAATGTGCCAACCCAGG	1740
Db	1586	CTTAAAGATGCTCTCTCCAGCCCTGTGTCTCAACCAATCTCCAAATTAATGTGCCAACCCAGG	1645
Oy	1741	GGCCCTGAGACCTCCCAACATTCATTCATTCCTTGTGCTGCGCAAGTGGCAATPAACGGGTGAT	1800
Db	1646	GGCCCTGAGACCTCCCAACATTCATTCATTCCTTGTGCTGCGCAAGTGGCAATPAACGGGTGAT	1705
Oy	1801	TGCC 1804	
Db	1706	TGCC 1709	
RESULT 8			
LOCUS	BD158376	1614 bp	DNA linear
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD158376		
VERSION	BD158376.1		
KEYWORDS	JP 2002191363-A/13219		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 1614)		
TITLE	Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.		
JOURNAL	Primer for synthesizing full-length cDNA and use thereof		
COMMENT	Patent: JP 2002191363-A 13219 09-JUL-2002; HELIX RESEARCH INSTITUTE		
OS	Homo sapiens (human)		
PN	JP 2002191363-A/13219		
PD	09-JUL-2002		
PF	28-JUL-2000 JP 2000280990		
PI	TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAOBU		
PI	SAITO,		
PI	YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,		
PI	KEIICHI NAGAI,TETSUJI OTSUKI		
PC	C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/10,		
PC	C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC		
Primer for synthesizing full-length cDNA and use thereof FH			
Location/Qualifiers			
FT	CDS	Location/Qualifiers	
1..1614			
/organism="Homo sapiens"			
/mol_type="Genomic DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Query Match	79.2%;	Score 1434.8;	DB 6;
Best Local Similarity	90.6%;	Pred. No. 0;	
Matches 1612;	Conservative 0;	Mismatches 2;	Indels 166;
Gaps 1;			
25	AGTACACAGCACCCTGAGTGTGTGACACACAGAGGACCCCTTCATGTTTAAAGGACCTCT	84	
Db	1	AGTACACAGCACCCTGAGTGTGTGACACACAGAGGACCCCTTCATGTTTAAAGGACCTCT	60
Oy	85	GGGCTCTAGAGAGCTGTGGGCGCGCCCTCTGTGGGAGACTCCCCCATCCGCGGGCGGAATG	144
Db	61	GGGCTCTAGAGAGCTGTGGGCGCGCCCTCTGTGGGAGACTCCCCCATCCGCGGGCGGAATG	120
Oy	145	GTCGCGGTGCGCTGTGCGAGTGTGTGTGCTGCTCCCTGTGATTCGTGGATGCAAGTGTGG	204

Db	121	GTCCGGGCGCGTCCGACAGTCTCTGCTCTCCCTGCTTTCCTGGGTGCAAAATGCTGG	180
OY	205	GTTCCTGGGATTCTGGAATTCGCGGGCGCTTGACACAGTAGCTGTGCGGCTCTTCGGGTGA	264
Db	181	GTTCCTGGGATTCTGGAATTCGCGGGCGCTTGACACAGTAGCTGTGTCGGCTTCCTGGGTGA	240
OY	265	GTCCGTCCGCGCGGGTGTGCCCCGGGAAAGGCTTAGAGCTCCGGGGTTCGGGGGCCCAAGGC	324
Db	241	GTCCGTCCGCGCGGGTGTGCCCCGGGAAAGGCTTAGAGCTCCGGGGTTCGGGGGCCCAAGGC	300
OY	325	ATTCCGGGCTTCAGATTGACGGGGATCCCGGATTCACCGCGCGCCCCCGGCGCTTCAACG	384
Db	301	ATTCCGGGCTTCAGATTGACGGGGATCCCGGATTCACCGCGCGCCCCCGGCGCTTCAACG	360
OY	385	ACGGGTCCAGACTGTGTGGGAAGAGTGCGGGACGGGTCTCTGAGGATCCGATCCT	444
Db	361	ACGGGTCCAGACTGTGTGGGAAGAGTGCGGGACGGGTCTCTGAGGATCCCGATCCT	420
OY	445	ACGAGCCAAAGTGTCAAGCTTTATAGGTGTGACCTTACATGTGACTTTACCTCAGTTT	504
Db	421	ACGAGCCAAAGTGTCAAGCTTTATAGGTGTGACCTTACATGTGACTTTACCTCAGTTT	480
OY	505	GTGATCCGTAATAATGACAATTTGGAAGCTACTTCAAGAGCTGTGTGAGAGATTAATG	564
Db	481	GTGATCCGTAATAATGACAATTTGGAAGCTACTTCAAGAGCTGTGTGAGAGATTAATG	540
OY	565	AAACAATGCTTGTAAAGCTCTTTGACAGAGGAGCCTCGGAAGCAGGGCCTGCGCGGAG	624
Db	541	AAACAATGCTTGTAAAGCTCTTTGACAGAGGAGCCTCGGAAGCAGGGCCTGCGCGGAG	600
OY	625	AGCACACTGCTGTACACAGGAGACACAGGACGATGAAACCCCTCGTAGCTGTGCTCT	684
Db	601	AGCACACTGCTGTACACAGGAGACACAGGACGATGAAACCCCTGTGAGCTGTGCTCT	660
OY	685	CAGTGGGATGCAAGCCCTCGGCTTTCAGCAACGCTGCGGAGGTGTGCTAACCGGGTCAAGGC	744
Db	661	CAGTGGGATGCAAGCCCTCGGCTTTCAGCAACGCTGCGGAGGTGTGCTAACCGGGTCAAGGC	720
OY	745	CAGAGCTCATATGTGATGATGAGACTCTGTTTGGCAGCCACAGGACACCCGACTTACCC	804
Db	721	CAGAGCTCATATGTGATGATGAGACTCTGTTTGGCAGCCACAGGACACCCGACTTACCC	780
OY	805	ACCAGACTTGATTCGCCCTGGGTGTGAGGAAGGCTTAACAAACAGAGGCGTGGGCAAGGA	864
Db	781	ACCAGACTTGATTCGCCCTGGGTGTGAGGAAGGCTTAACAAACAGAGGCGTGGGCAAGGA	840
OY	865	GGCATTCAGAGGCTTTGGGGGCAAAAGGGAGCTGTGAACCAACCCCTCAAGGGGCAAGC	924
Db	841	GGCATTCAGAGGCTTTGGGGGCAAAAGGGAGCTGTGAACCAACCCCTCAAGGGGCAAGC	900
OY	925	CCCCACCTTCACACCAAGGAAGAAACAATACAGACCCATCAGCCACACCCCGTCTTA	984
Db	901	CCCCACCTTCACACCAAGGAAGAAACAATACAGACCCATCAGCCACACCCCGTCTTA	960
OY	985	CTGTGATGATGCTCGCTGTTTGGCTCCGATCTGAAGGCGCAAGCTTCGGGGCCCGCGGAT	1044
Db	961	CTGTGATGATGCTCGCTGTTTGGCTCCGATCTGAAGGCGCAAGCTTCGGGGCCCGCGGAT	1020
OY	1045	GGCGAAGGGGATATCCGCAAAAGCTCGGTGCTCTTGTGACGGCAACACACTACCCGAG	1104
Db	1021	GGCGAAGGGGATATCCGCAAAAGCTCGGTGCTCTTGTGACGGCAACACACTACCCGAG	1080
OY	1105	GGGTAGCCACTCGCGCCCGCCCAAGGAGAGCAACACTCGAGCCATTACCCAGCTGTGTC	1164
Db	1081	GGGTAGCCACTCGCGCCCGCCCAAGGAGAG-----	1109
OY	1165	CTCCAGAAGAGCCGCGGGCCAGCGGACAGACTCCAGAAGTTATATGAGGTGAGTTACA	1224
Db	1140	-----	1109
OY	1225	CTCTTACAGCCCCCTGTAAGCGGGGACTTTTCCATTCCCTCACCCACTGATGTCCCGAG	1284
Db	1100	-----	1109





```

QY 1105 GGGTAGCCTCGCCCGCCGAGGAGCACCAGTCCAGCATTCACCGAGTGGTCC 1164
DB 1081 GGGTAGCCTCGCCCGCCGAGGAGG----- 1109
QY 1165 CTCGACAGCAGAGCCGGGGCAGCGGAGACTCCAGAGATTATCTATGAGTGGTTCACA 1224
DB 1110 ----- 1109
QY 1225 CTCTTACGCCCCCTGAAAGCGGGGAGCTTTCCCATTCCTCCACCACTGAATGTCCGAG 1284
DB 1110 ----- 1109
QY 1285 CACTGTCATCAGCCAGCCAGTGCCTCCCAACAAATGGGCTTCAGAGCTTC 1344
DB 1110 -----CACCAGTGCCTCCCAACAAATGGGCTTCAGAGCTTCAGAGCTTC 1154
QY 1345 CAGGTACAGGGGTGACCTTCCGAGAGCCCTGTGTGATCTTCCAGGCTCGCTAGTTAGCAT 1404
DB 1155 CAGGTACAGGGGTGACCTTCCGAGAGCCCTGTGTGATCTTCCAGGCTCGCTAGTTAGCAT 1214
QY 1405 TTCAAGTCCATCTTACCCAGAGAGTGGGGCCACCCAGAACCAAGCCCTTGGAA 1464
DB 1215 TTCAAGTCCATCTTACCCAGAGAGTGGGGCCACCCAGAACCAAGCCCTTGGAA 1274
QY 1465 ATGATACCTTTTATCAGGGTGGCTATGGGGCCAGGGCCAGGTATGGCCCTTGGCA 1524
DB 1275 ATGATACCTTTTATCAGGGTGGCTATGGGGCCAGGGCCAGGTATGGCCCTTGGCA 1334
QY 1525 GGGTAGAGGAGCATTCATCAACCCAGGAAACCCAGATTAAGAGACCCCTGTGGGGC 1584
DB 1335 GGGTAGAGGAGCATTCATCAACCCAGGAAACCCAGATTAAGAGACCCCTGTGGGGC 1394
QY 1585 AGACAGACATAGCAGGGGTGGGGCTGCTCCCTTTATCTGACAACTCTAGTGCATTC 1644
DB 1395 AGACAGACATAGCAGGGGTGGGGCTGCTCCCTTTATCTGACAACTCTAGTGCATTC 1454
QY 1645 TTGCGCTTTTCTCCGATGCGGATTTGGGGCCACCTTAAGATGCTCTTCAGAGCC 1704
DB 1455 TTGCGCTTTTCTCCGATGCGGATTTGGGGCCACCTTAAGATGCTCTTCAGAGCC 1514
QY 1705 TGTCTCAACCATCTCCAAATTAAGTGCACACCGAGGGCTGTGACCTCCCAATCATTC 1764
DB 1515 TGTCTCAACCATCTCCAAATTAAGTGCACACCGAGGGCTGTGACCTCCCAATCATTC 1574
QY 1765 ATTGTCTTGTGCTGCAAGTGGGAAATAAACGGGTGATGGCC 1804
DB 1575 ATTGTCTTGTGCTGCAAGTGGGAAATAAACGGGTGATGGCC 1614

RESULT 10
AK027741 1614 bp mRNA linear PRI 30-JAN-2004
LOCUS Homo sapiens cDNA FLJ14835 fis, clone OVARC1001369.
ACCESSION AK027741
VERSION AK027741.1 GI:14042642
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuna, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanihara, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Miyakawa, K., Fujimori, K., Tanai, H., Kinata, M., Watanabe, M.,
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,

```

```

Yosida, M., Horuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Mori, S., Nomiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
JOURNAL
TITILE
REFERENCE
AUTHORS
JOURNAL
TITILE
COMMENT
FEATURES
SOURCE
CDS
ORIGIN
misc_feature
Query Match 79.2%; Score 1434.8; DB 9; Length 1614;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
25 ATGACACAGCAGCCTGAGTAGAGGAGCAGCAGAGCCTTCATGTTTGGGAGCTTCT 84

```

Db 1 AGTACACAGCAGCTGATGATGAGCAACAGAGACCTCTCATGTTTAAAGGACCTCT 60  
 Qy 85 GGGGCTCAGAGACGTGAGCGCCCGCCCTTGGGCGAGCTCCCGCATCCGCGGCGCAATG 144  
 Db 61 GGGGCTCAGAGACGTGAGCGCCCGCCCTTGGGCGAGCTCCCGCATCCGCGGCGCAATG 120  
 Qy 145 GTCCGGGTCGCGTCCGAGTGTCTGCTGCTGCTCCCTGGTTGCTGGGTTCAAAAGTCTGG 204  
 Db 121 GTCCGGGTCGCGTCCGAGTGTCTGCTGCTGCTCCCTGGTTGCTGGGTTCAAAAGTCTGG 180  
 Qy 205 GTTCTGGGTTTGTGATTCGCGGCGCTTCAACAGTACCTGTGCGGCTCTCTCGGTTGA 264  
 Db 181 GTTCTGGGTTTGTGATTCGCGGCGCTTCAACAGTACCTGTGCGGCTCTCTCGGTTGA 240  
 Qy 265 GTCCGTCGCGCGCGGTCGCGGCGCTAGGCTGCGCGGCGGTCGCGGCGCCCGAGGC 324  
 Db 241 GTCCGTCGCGCGCGGTCGCGGCGCTAGGCTGCGCGGCGGTCGCGGCGCCCGAGGC 300  
 Qy 325 ATTCCGGGCTGAGATTGACGGGGAATCCCGATGACCGCGCGCCCGCGCTTCAACCG 384  
 Db 301 ATTCCGGGCTGAGATTGACGGGGAATCCCGATGACCGCGCGCCCGCGCTTCAACCG 360  
 Qy 385 ACGGATCCAGACCTGTGAGGAAAGAGTCCGGGACGGGTCCTTGAAGATCCCGATGCT 444  
 Db 361 ACGGATCCAGACCTGTGAGGAAAGAGTCCGGGACGGGTCCTTGAAGATCCCGATGCT 420  
 Qy 445 ACGAGCCAAAGATGCTCAGCTTTATAGGTGTGACTACACATGTAATCTTCACTCACTGTTT 504  
 Db 421 ACGAGCCAAAGATGCTCAGCTTTATAGGTGTGACTACACATGTAATCTTCACTCACTGTTT 480  
 Qy 505 GTGATCCGTAATGAGCAAAATTCGAAGCTACTTCAAGTGTGTTGAGAGATTTAAATG 564  
 Db 481 GTGATCCGTAATGAGCAAAATTCGAAGCTACTTCAAGTGTGTTGAGAGATTTAAATG 540  
 Qy 565 AAACATATGCTTTGTAAGCTTTTGAAGAGGAGCTTGGGAAAGACAGGCTTGGCGGAG 624  
 Db 541 AAACATATGCTTTGTAAGCTTTTGAAGAGGAGCTTGGGAAAGACAGGCTTGGCGGAG 600  
 Qy 625 AGCAACCTGCTGTACACAGGAGCAACAGGAGCAATGAAGACCCCGTGGAGCTGGCGT 684  
 Db 601 AGCAACCTGCTGTACACAGGAGCAACAGGAGCAATGAAGACCCCGTGGAGCTGGCGT 660  
 Qy 685 CAGTGGATGAGACCTTCGCGCTTCAAGACCGCTGCGAGGTTGCTACCGGTTCAAGGC 744  
 Db 661 CAGTGGATGAGACCTTCGCGCTTCAAGACCGCTGCGAGGTTGCTACCGGTTCAAGGC 720  
 Qy 745 CAGGAGTCTATATGTGATGAGACTCTGTTTGGACGCCAGGACCCCGCTTACCC 804  
 Db 721 CAGGAGTCTATATGTGATGAGACTCTGTTTGGACGCCAGGACCCCGCTTACCC 780  
 Qy 805 ACCGGAATTTGATTCGCGCTGCTGGTGGAGAGGCTTAAAGACCAAGGCTTGGGCAAGA 864  
 Db 781 ACCGGAATTTGATTCGCGCTGCTGGTGGAGAGGCTTAAAGACCAAGGCTTGGGCAAGA 840  
 Qy 865 GGCATCGAAGGCTTGGGAGGCAAGAGGAGCTGTGAGACACCCCTTCAAGGAGGAGCAGC 924  
 Db 841 GGCATCGAAGGCTTGGGAGGCAAGAGGAGCTGTGAGACACCCCTTCAAGGAGGAGCAGC 900  
 Qy 925 CCCCACTCTCAACCAAGAGAGAGCAAAATATACAGACCCATCAGCCACCCCGTCTTA 984  
 Db 901 CCCCACTCTCAACCAAGAGAGAGCAAAATATACAGACCCATCAGCCACCCCGTCTTA 960  
 Qy 985 CTGATGATGATGCTGCTTTGGCTCCCGATCTGAAGGCGCAGCTTGGGAGGCGCGGAT 1044  
 Db 961 CTGATGATGATGCTGCTTTGGCTCCCGATCTGAAGGCGCAGCTTGGGAGGCGCGGAT 1020  
 Qy 1045 GGCAGAGGAGGATGCGCAAGAGCTCCGATCTCTTGTGAGAGGACCACTAACCCCGAG 1104  
 Db 1021 GGCAGAGGAGGATGCGCAAGAGCTCCGATCTCTTGTGAGAGGACCACTAACCCCGAG 1080  
 Qy 1105 GGGTAGGCACTGCGCCCGCCCGAGGAGGACCACTGCGAGGCACTTCAACCGAGTGTCC 1164  
 Db 1081 GGGTAGGCACTGCGCCCGCCCGAGGAGG----- 1109

Qy 1165 CTCCAAGACAGAGCCGGGCGCAGCGGACAGACTCCAGAAATTATGATGATGAGTTACA 1224  
 Db 1110 ----- 1109  
 Qy 1225 CTCTTCAGCGCCCTGAAAGCGGGACTTTCCTATTCCTCAGCCACTGAATGTCCGAG 1284  
 Db 1110 ----- 1109  
 Qy 1285 CACTGATATCCAGCCACCATGTCCTCCCGACCAAAATGGGCTCAGATCTCAGGCTTC 1344  
 Db 1110 -----CACCAAGTCCCGCCACCAAAATGGGCTCAGATCTCAGGCTTC 1154  
 Qy 1345 CACGTCAAGGAGTGAATCTCCGAGGCGCCCTGATGACTTCCAGGCTGCTCAGTTAGCAT 1404  
 Db 1155 CACGTCAAGGAGTGAATCTCCGAGGCGCCCTGATGACTTCCAGGCTGCTCAGTTAGCAT 1214  
 Qy 1405 TTCACTGCTCATCTAACCCACGACGAGTGGGCGCACCCAGAAACCAAGGCCCTTGGAA 1464  
 Db 1215 TTCACTGCTCATCTAACCCACGACGAGTGGGCGCACCCAGAAACCAAGGCCCTTGGAA 1274  
 Qy 1465 ATGATACCTTTTATCAGGCTTGTCTATGAGGCGCACCGGACAGGATATGCCCCCTTGCA 1524  
 Db 1275 ATGATACCTTTTATCAGGCTTGTCTATGAGGCGCACCGGACAGGATATGCCCCCTTGCA 1334  
 Qy 1525 GGGTAGAGGACATTCATCACCCAGGAAACCCAGATTTAAAGAGGCCCTGTGGGGGC 1584  
 Db 1335 GGGTAGAGGACATTCATCACCCAGGAAACCCAGATTTAAAGAGGCCCTGTGGGGGC 1394  
 Qy 1585 AGACAGATATGACAGGAGTGGGCAAGTGTGCTCTCTTATCTGACATCTTATGATGATTC 1644  
 Db 1395 AGACAGATATGACAGGAGTGGGCAAGTGTGCTCTCTTATCTGACATCTTATGATGATTC 1454  
 Qy 1645 TTGCTTTTCTCCGATTTGGGAGATTGGGGGCACTCTTAAAGATGCTCTTCCAGGCC 1704  
 Db 1455 TTGCTTTTCTCCGATTTGGGAGATTGGGGGCACTCTTAAAGATGCTCTTCCAGGCC 1514  
 Qy 1705 TGTCTCAACCAATCTCCCAATTTAGTGGCAACCCAGGAGCTGGGACCTCCCAATCATCC 1764  
 Db 1515 TGTCTCAACCAATCTCCCAATTTAGTGGCAACCCAGGAGCTGGGACCTCCCAATCATCC 1574  
 Qy 1765 ATTGTCTTGTCCCAAGTGGCAATTTAAAGCGGCTGATTTCC 1804  
 Db 1575 ATTGTCTTGTCCCAAGTGGCAATTTAAAGCGGCTGATTTCC 1614

RESULT 11  
 BC022092  
 LOCUS  
 DEFINITION Homo sapiens hypothetical protein FLJ14827, mRNA (cdna clone  
 MGC:4570 IMAGE:3029785), complete cds.  
 ACCESSION BC022092  
 VERSION BC022092.2 GI:33870282  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1407)  
 AUTHORS  
 Klausner R.D., Collins E.A., Grouse L.H., Derge J.G.,  
 Strausberg R.L., Feingold F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L.,  
 Scheetz T.E., Brownstein M.J., Ueda T.B., Toshitake S.,  
 Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,  
 Abramson R.D., Mullen S.J., Bosak S.A., McEwan P.J.,  
 McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,  
 Worsley K.C., Hale S., Garcia A.M., Gay L.J., Huily S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S.,  
 Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Bretter, J.D., Krzywicki, M.I., Skalska, U., Snailus, D.E.,  
Schmechel, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1407)  
Strausberg, R.  
Direct Submission  
Submitted (22-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Aug 19, 2003 this sequence version replaced gi:18314382.  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
Adnan Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha  
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whitting  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ULNL at: <http://image.jhmi.gov>  
Series: IRAL Plate: 10 Row: h Column: 14  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 14249569.  
Location/Qualifiers  
1. 1407  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:4570 IMAGE:3029785"  
/tissue\_type="Ovary, adenocarcinoma"  
/clone\_id="NIH\_MGC\_9"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
1. 1407  
/gene="FLJ14827"  
/db\_xref="LocusID:84934"  
224. 1033  
/gene="FLJ14827"  
/codon\_start=1  
/product="hypothetical protein FLJ14827"  
/protein\_id="AAH2092.1"  
/db\_xref="GI:18314383"  
/translation="MKTPELVASGMQTLGIQHRCRGKRVKARTSYVDETLFGSPAG  
ISHPSPYDPSFVWEKARTRGVGEKAKAGKSCPTPSRSTPLTPRKNTYRP  
LRTHPAGPSKTERGPAADSOIKSMGLHSRPLKRGSLSHLTHVNPSTGHATLSAP  
HTNPQDLRPSSTGVTFRSPLVTSRAVSISVPTFRRGATQKRPWK"

ORIGIN  
Query Match 75.7%; Score 1371; DB 9; Length 1407;  
Best Local Similarity 100.0%; Pred. No. 5.1e-300;  
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

gene  
CDS  
436 CCGATGCTTACGACCGAAGATGCTCAGCTTTATAGGTGACCTACACATGTGACTTAC 495  
Db 1 CCGATGCTTACGACCGAAGATGCTCAGCTTTATAGGTGACCTACACATGTGACTTAC 60  
Qy 496 CTCAGTTTGTGATCCGTAATAATGACAAATTCAGACTACTTACAGTGTCTTTGAGAG 555  
Db 61 CTCAGTTTGTGATCCGTAATAATGACAAATTCAGACTACTTACAGTGTCTTTGAGAG 120  
Qy 556 GATTAATGAAACAATGCTTTGTAAGCTCTTGACAGAGGAGCCTCGAAGACAGGACT 615

Db 121 GATTAATGAAACAATGCTTTGTAAGCTCTTGACAGAGGAGCCTCGAAGACAGGACT 180  
Qy 616 GGGCGGACAGACACCTGTGTGACACAGAGACCAAGAGATGAAGACCCCGTGA 675  
Db 181 GGGCGGACAGACACCTGTGTGACACAGAGACCAAGAGATGAAGACCCCGTGA 240  
Qy 676 GCTGCGGTGAGAGGATGAGAGACCTGTGACACCGGTGCGAGGTGACTACCG 735  
Db 241 GCTGCGGTGAGAGGATGAGAGACCTGTGACACCGGTGCGAGGTGACTACCG 300  
Qy 736 GGTCAAGGCGAGACGTCATATGTGATGAGACTCTGTGTCAGACCCGACAGCCG 795  
Db 301 GGTCAAGGCGAGACGTCATATGTGATGAGACTCTGTGTCAGACCCGACAGCCG 360  
Qy 796 GCTTACCCGACCGGACTTGTGATCCGCTGGGTGAGAGGCTTAACAGACAGAGCGT 855  
Db 361 GCTTACCCGACCGGACTTGTGATCCGCTGGGTGAGAGGCTTAACAGACAGAGCGT 420  
Qy 856 GGGCAAGAGAGCATGCGAAGGCTTGGGGGCAAGGAGGCTGTGAGACACCCCTCAAG 915  
Db 421 GGGCAAGAGAGCATGCGAAGGCTTGGGGGCAAGGAGGCTGTGAGACACCCCTCAAG 480  
Qy 916 GGGCAAGACCCCACTCAACCAAGAAAGAAACAATAACAGACCCATCAAGCAC 975  
Db 481 GGGCAAGACCCCACTCAACCAAGAAAGAAACAATAACAGACCCATCAAGCAC 540  
Qy 976 CCGGTCTTACGTGATGAGTGGCTGTGTTGGCTCCCATCTGAAGGGGCCGCTGGGGG 1035  
Db 541 CCGGTCTTACGTGATGAGTGGCTGTGTTGGCTCCCATCTGAAGGGGCCGCTGGGGG 600  
Qy 1036 CCGCGGATGGCGAAGGGGGATGCGCAAGTCCGCTCTCTGTGAGCGCACAC 1095  
Db 601 CCGCGGATGGCGAAGGGGGATGCGCAAGTCCGCTCTCTGTGAGCGCACAC 660  
Qy 1096 TACCCCGAGGGGTAGCCACTGCTGCTCCGCGCCAGAGAGCCACATGAGCAATTCACC 1155  
Db 661 TACCCCGAGGGGTAGCCACTGCTGCTCCGCGCCAGAGAGCCACATGAGCAATTCACC 720  
Qy 1156 AGCTGTGCTCTCAACAACAGAGCGGGGCGAGGGGAGACTCCAGAGATCTATGAGG 1215  
Db 721 AGCTGTGCTCTCAACAACAGAGCGGGGCGAGGGGAGACTCCAGAGATCTATGAGG 780  
Qy 1216 TGGATTACACTCTTCAAGCGCCCTGAGAGCGGGGACTTCCATTTCCCTACCCACTGAA 1275  
Db 781 TGGATTACACTCTTCAAGCGCCCTGAGAGCGGGGACTTCCATTTCCCTACCCACTGAA 840  
Qy 1276 TGTCCCAAGACGTGTATCATCAGCCACAGTGTCCCTCCACAAATGGGCTCAGATCT 1335  
Db 841 TGTCCCAAGACGTGTATCATCAGCCACAGTGTCCCTCCACAAATGGGCTCAGATCT 900  
Qy 1336 CAGGCTTCCACGTCAGGGGTGACTTCCGGAAGCCCTGTGTGACTTCCAGGGCTGCTC 1395  
Db 901 CAGGCTTCCACGTCAGGGGTGACTTCCGGAAGCCCTGTGTGACTTCCAGGGCTGCTC 960  
Qy 1396 AGTTAGATTTCAGTGCATCTACCCCAAGAGAGTGGGGCCACCAAGAACCAAGGCC 1455  
Db 961 AGTTAGATTTCAGTGCATCTACCCCAAGAGAGTGGGGCCACCAAGAACCAAGGCC 1020  
Qy 1456 CCCTTGAATGATATCTTTTATCAGAGGTGCTTATGGGGCCAGCGGACAGGTATGCG 1515  
Db 1021 CCCTTGAATGATATCTTTTATCAGAGGTGCTTATGGGGCCAGCGGACAGGTATGCG 1080  
Qy 1516 CCCTTGCAGAGGTGAGAGACATTCATCAACCCAGGGAAACCCAGATTTAAAGGCC 1575  
Db 1081 CCCTTGCAGAGGTGAGAGACATTCATCAACCCAGGGAAACCCAGATTTAAAGGCC 1140  
Qy 1576 TGTGGGGGAGAGACATATGACAGGGGTGGGAGTGTGCTCTTATCTGACATCTCT 1635  
Db 1141 TGTGGGGGAGAGACATATGACAGGGGTGGGAGTGTGCTCTTATCTGACATCTCT 1200  
Qy 1636 AGTGAATCTTGTGCTTTTCTCCGATTTGGGAGCCACTTAAGATGCTCT 1695

Db 1201 AGTCATCTTCTGCTTTTCTCCCGATTGCGGAGTTGGGGGCGACCTCTAAGATGCTCT 1260

Qy 1696 CTCGAGCCCTGCTCTAACCATCTCCAAATTAATGTCGCAACCGAGGGGCGCTGACCTGCC 1755

Db 1261 CTCGAGCCCTGCTCTAACCATCTCCAAATTAATGTCGCAACCGAGGGGCGCTGACCTGCC 1320

Qy 1756 ACATCATCATCTGCTCTGCTGCGCAAGTGGCAATTAACGCGCGATTTGCCAA 1806

Db 1321 ACATCATCATCTGCTCTGCTGCGCAAGTGGCAATTAACGCGCGATTTGCCAA 1371

RESULT 12  
AC089999/C  
LOCUS  
DEFINITION Homo sapiens 12 BAC RP11-545P7 (Roswell Park Cancer Institute Human  
BAC library) complete sequence.  
ACCESSION AC089999 AC021161  
VERSION AC089999.20 GI:21617628  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 199522)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Albriock,S.L., Amaralungue,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieve,M., Brown,B., Brown,M., Bryant,N.P.,  
Bunay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chit,D., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
DeLaney,K.R., Delgado,O., Demn,A.U., Ding,Y., Dim,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,J.,  
Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,  
Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,  
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlik,P.,  
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,  
Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huliy,S.,  
Hume,J., Ioshikes,I., Jackson,L.E., Jacobson,B., Jia,Y.,  
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,  
King,L., Korrah,V., Kovar,C., Kravovic,J., Kureishi,A., Landry,N.,  
Lee,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichargue,O.,  
Lieu,C., Liu,J., Liu,W., Loulesged,H., Lozada,R.J., Lu,X.,  
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P.,  
Mandrel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,  
Mashiny,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S.,  
Metzke,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,  
Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,  
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguh,M., Okunnu,G.,  
Orgunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peterson,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G.,  
Scherer,S., Scott,G., Shen,H., Shim,C., Shooshari,N., Sisson,I.,  
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Umanai,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Warren,R., Wleczek,R., Woodson,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,  
Weinstock,G. and Gibbs,R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 199522)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (08-FEB-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE  
AUTHORS Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 199522)  
Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 199522)  
Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 199522)  
Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
6 (bases 1 to 199522)  
Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (23-OCT-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 27, 2002 this sequence version replaced gi:21490131.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

COMMENT  
AUTHORS  
TITLE  
JOURNAL

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>  
ml.

FEATURES  
source  
Location/Qualifiers  
1..199522  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-545P7"  
/complement(25..342)  
/rpt\_family="AluSx"  
complement(52..2005)  
repeat\_region  
misc\_feature

Query Match 47.2%; Score 854; DB 9; Length 199522;  
Best Local Similarity 100.0%; Pred. No. 6.8e-183;  
Matches 854; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 958 CAGACCATGACGACACCCGCTTACTGTGATGAGTGCCTGTTGGCTCCGATCTGA 1017

/note="overlaps bases 1..1954 of clone AC127004"  
/function="clone overlap"  
814..1156  
STs  
repeat\_region /strand.name="SHGC-148490"  
complement(943..1150)  
/rpt\_family="MIR"  
repeat\_region complement(1487..1598)  
/rpt\_family="MIR"  
2041..2731  
repeat\_region /rpt\_family="L2"  
complement(5385..5699)  
/rpt\_family="AluY"  
repeat\_region complement(5720..6033)  
/rpt\_family="AluB"  
5755..5967  
STs  
repeat\_region /strand.name="SHGC-67307"  
complement(7231..7531)  
/rpt\_family="AluSg"  
repeat\_region complement(7633..7927)  
/rpt\_family="AluSx"  
repeat\_region complement(7986..8413)  
/rpt\_family="L2"  
repeat\_region complement(8408..8611)  
/rpt\_family="MIR"  
repeat\_region complement(9109..9166)  
/rpt\_family="L2"  
9207..9256  
repeat\_region /rpt\_family="(CA)n"  
9666..10270  
/rpt\_family="AluSg"  
repeat\_region complement(11877..12165)  
/rpt\_family="AluB"  
12427..12500  
/rpt\_family="AluSx"  
13357..13658  
/rpt\_family="AluSx"  
repeat\_region complement(14062..14304)  
/rpt\_family="MIR"  
repeat\_region complement(14379..14482)  
/rpt\_family="AluSg/x"  
14642..14738  
/rpt\_family="MER102"  
15557..15798  
STs  
repeat\_region /strand.name="Sts-AA001615"  
17094..17319  
/rpt\_family="MIR"  
repeat\_region complement(18276..18388)  
/rpt\_family="MIR"  
19450..19813  
repeat\_region /rpt\_family="MLT1B"  
20261..20366  
/rpt\_family="AluSx"  
repeat\_region 20367..20547  
/rpt\_family="MLT1B"  
20548..20733  
repeat\_region /rpt\_family="AluSx"  
complement(21056..21223)  
/rpt\_family="L2"  
21224..21514  
/rpt\_family="AluSx"  
repeat\_region complement(21515..21595)  
/rpt\_family="L2"  
22244..22372  
repeat\_region /rpt\_family="MIR"  
complement(22394..22690)  
/rpt\_family="AluB"  
22708..22777

Db 59708 CAGACCATGACGACACCCGCTTACTGTGATGAGTGCCTGTTGGCTCCGATCTGA 59649  
QY 1018 AGCGGACGCTTGGGGGCCCCGGGATGAGGAGGGGATGCGCAAAAGCTCCGTGCT 1077  
Db 59648 AGCGGACGCTTGGGGGCCCCGGGATGAGGAGGGGATGCGCAAAAGCTCCGTGCT 59589  
QY 1078 CTGTGGACGCAACCACTTACCCCGAGGGGTAGCACTCGCCCCGCCCGGAGGAGGAC 1137  
Db 59588 CTGTGGACGCAACCACTTACCCCGAGGGGTAGCACTCGCCCCGCCCGGAGGAGGAC 59529  
QY 1138 ACTGCGAGCATTCACCACTGCTGCTCCCAAGACAGAGCCGAGGCGAGGAGCTC 1197  
Db 59528 ACTGCGAGCATTCACCACTGCTGCTCCCAAGACAGAGCCGAGGCGAGGAGCTC 59469  
QY 1198 CGAAGATTATCATGGGAGGGGTTACACTTACGCGCCCCGTAAGCGGGAGCTTCCCA 1257  
Db 59468 CGAAGATTATCATGGGAGGGGTTACACTTACGCGCCCCGTAAGCGGGAGCTTCCCA 59409  
QY 1258 TTCCCTCACCACTGATATGTCCTCCAGCACTGGTCAATCCAGCAGAGTCCGCCCA 1317  
Db 59408 TTCCCTCACCACTGATATGTCCTCCAGCACTGGTCAATCCAGCAGAGTCCGCCCA 59349  
QY 1318 AAATGGGCTCAGGATCTCAGGCTTCCAGCTCAGGAGTGAAGCTTCCGAGGCTGCT 1377  
Db 59348 AAATGGGCTCAGGATCTCAGGCTTCCAGCTCAGGAGTGAAGCTTCCGAGGCTGCT 59289  
QY 1378 GACTTCAGGGGCTGCTCAGTATGAGCAATTCAGGAGCAATCCAGAGAGTGGGGC 1437  
Db 59288 GACTTCAGGGGCTGCTCAGTATGAGCAATTCAGGAGCAATCCAGAGAGTGGGGC 59229  
QY 1438 CACCCAGAAACCAAGGCCCTTGAATGATCTTTTCATCAGGGTTCCTATGGGGC 1497  
Db 59228 CACCCAGAAACCAAGGCCCTTGAATGATCTTTTCATCAGGGTTCCTATGGGGC 59169  
QY 1498 CACGGGACAGGATGAGGCCCCCTTGAAGTGAAGAGGATTCATCAGGAGGAGGAG 1557  
Db 59168 CACGGGACAGGATGAGGCCCCCTTGAAGTGAAGAGGATTCATCAGGAGGAGGAG 59109  
QY 1558 AGGTATTAAGAGAGCCCTGAGGGGAGAGCAATGAGAGGGGAGGAGTGCCTCC 1617  
Db 59108 AGGTATTAAGAGAGCCCTGAGGGGAGAGCAATGAGAGGGGAGGAGTGCCTCC 59049  
QY 1618 TTATCTGACAAATCTAGTGCATTCCTGCTTTTTCCTCCGATTCGAGATTCGGGGC 1677  
Db 59048 TTATCTGACAAATCTAGTGCATTCCTGCTTTTTCCTCCGATTCGAGATTCGGGGC 58989  
QY 1678 CACCTTAAGATGCTCTCTCCAGCCCTGTCACACCATATCTCAAAATTAATGTCAC 1737  
Db 58988 CACCTTAAGATGCTCTCTCCAGCCCTGTCACACCATATCTCAAAATTAATGTCAC 58929  
QY 1738 AGGGGCTGCGACCTCCCATATCATCATGCTGCTGTCAGAGTGAATTAACGGCGT 1797  
Db 58928 AGGGGCTGCGACCTCCCATATCATCATGCTGCTGTCAGAGTGAATTAACGGCGT 58869  
QY 1798 GATTGCCAACCTGG 1811  
Db 58868 GATTGCCAACCTGG 58855

RESULT 13  
BD149092 784 bp DNA linear PAT 17-JAN-2003  
LOCUS BD149092  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD149092  
VERSION BD149092.1 GI:27854850  
KEYWORDS JP 2002191363-A/3935.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 784)  
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,T.,  
AUTHORS



TITLE  
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
Primer for synthesizing full-length cDNA and use thereof  
Patent: JP 2002191363-A 3935 09-JUL-2002;  
JOURNAL  
HELIX RESEARCH INSTITUTE  
COMMENT  
OS Homo sapiens (human)  
PN JP 2002191363-A/3935  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI SAITO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC  
10, C12P21/02, C12O1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers  
FT source 1.784  
Location/Qualifiers  
FT source 1.784  
Location/Qualifiers  
1.784  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 36.5%; Score 661.4; DB 6; Length 784;  
Best Local Similarity 98.7%; Pred. No. 3.7e-139;  
Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 191 GTGCNAAAGTGTGCTGGGTTCTGGGTTCTGATTCGCGGGCCGTTCAACAGTACCTGTGCC 250  
DB 1 GTGCNAAAGTGTGCTGGGTTCTGGGTTCTGATTCGCGGGCCGTTCAACAGTACCTGTGCC 60  
QY 251 GGCCTCTGGGATGATCGGTCCGCGCGCGGATGCCCGGGACCGGCTAGGCTGCCGGGGGT 310  
DB 61 GGCCTCTGGGATGATCGGTCCGCGCGCGGATGCCCGGGACCGGCTAGGCTGCCGGGGGT 120  
QY 311 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 370  
DB 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180  
QY 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180  
DB 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180  
QY 371 CCGGCGCCTCAACGAGCGGATCCAGACTGTGGGAAAGGTCGGGGACCGGGTCCCTGA 430  
DB 181 CCGGCGCCTCAACGAGCGGATCCAGACTGTGGGAAAGGTCGGGGACCGGGTCCCTGA 240  
QY 431 GGATCCCGATGCTTACGAGCGCAAGATGCTCAAGCTTTATAGGTGACCTTACATGTGAC 490  
DB 241 GGATCCCGATGCTTACGAGCGCAAGATGCTCAAGCTTTATAGGTGACCTTACATGTGAC 300  
QY 491 TTCACCTCAAGTTTGTGATCCGTAAATGAGCAAAATTCGAAAGCTTACATGCTGTT 550  
DB 301 TTCACCTCAAGTTTGTGATCCGTAAATGAGCAAAATTCGAAAGCTTACATGCTGTT 360  
QY 551 GAGAGATTAATGAAACAATGCTTTGTAAGTCTTTTCAAGAGGAGGACCTTCGGAAGAG 610  
DB 361 GAGAGATTAATGAAACAATGCTTTGTAAGTCTTTTCAAGAGGAGGACCTTCGGAAGAG 420  
QY 611 GGCTGGCGCGGACAGCAACCTGCTGTCAACAGGAGCAACAGGAGATGAGAAGAGAGAG 670  
DB 421 GGCTGGCGCGGACAGCAACCTGCTGTCAACAGGAGCAACAGGAGATGAGAAGAGAGAG 480  
QY 671 GTGAGAGTGGCGGTCAAGTGGAGTGAACCTTCGAGCTTTCAGACCGGCTGCGAGAGTGC 730  
DB 481 GTGAGAGTGGCGGTCAAGTGGAGTGAACCTTCGAGCTTTCAGACCGGCTGCGAGAGTGC 540  
QY 731 TACCGGGTCAAGGCGCAGGACGTCAATGATGATG-AGAAGTCTGTTTGGCAGCCGAGAG 789  
DB 541 TACCGGGTCAAGGCGCAGGACGTCAATGATGATG-AGAAGTCTGTTTGGCAGCCGAGAG 600  
QY 790 CACCGGCTTACCCGAGCGGACCTTCGATCCGCTGGGTGGAGAGGCTTACAGAACGAG 849

DB 601 CACCGGCTTACCCGAGCGGACCTTCGATTCGCGGTGGAGAGGCTTACAGAACGAG 660  
QY 850 AGGCGTGGGCAAGGA-GGCATGGAAGGCTTGGGG 884  
DB 661 AGGCGTGGGCAAGGAAGGATTCAGAGGCTTGGGG 696

RESULT 14  
AX869030 784 bp DNA linear PAT 17-DEC-2003  
LOCUS  
DEFINITION Sequence 3935 from Patent EP1074617.  
ACCESSION AX869030  
VERSION AX869030.1 GI:40023893  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE  
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
Primer for synthesizing full-length cDNA and their use  
Patent: EP 1074617-A 3935 07-FEB-2001;  
JOURNAL Research Association for Biotechnology (JRP)  
FEATURES  
source 1.784  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 36.5%; Score 661.4; DB 6; Length 784;  
Best Local Similarity 98.7%; Pred. No. 3.7e-139;  
Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 191 GTGCNAAAGTGTGCTGGGTTCTGGGTTCTGATTCGCGGGCCGTTCAACAGTACCTGTGCC 250  
DB 1 GTGCNAAAGTGTGCTGGGTTCTGGGTTCTGATTCGCGGGCCGTTCAACAGTACCTGTGCC 60  
QY 251 GGCCTCTGGGATGATCGGTCCGCGCGCGGATGCCCGGGACCGGCTAGGCTGCCGGGGGT 310  
DB 61 GGCCTCTGGGATGATCGGTCCGCGCGCGGATGCCCGGGACCGGCTAGGCTGCCGGGGGT 120  
QY 311 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 370  
DB 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180  
QY 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180  
DB 121 CCGGGGCCCCAGGCAATTCGCGGCTGCAATTTGACGCGGATCCCGGATGACCGCGCGCC 180  
QY 371 CCGGCGCCTCAACGAGCGGATCCAGACTGTGGGAAAGGTCGGGGACCGGGTCCCTGA 430  
DB 181 CCGGCGCCTCAACGAGCGGATCCAGACTGTGGGAAAGGTCGGGGACCGGGTCCCTGA 240  
QY 431 GGATCCCGATGCTTACGAGCGCAAGATGCTCAAGCTTTATAGGTGACCTTACATGTGAC 490  
DB 241 GGATCCCGATGCTTACGAGCGCAAGATGCTCAAGCTTTATAGGTGACCTTACATGTGAC 300  
QY 491 TTCACCTCAAGTTTGTGATCCGTAAATGAGCAAAATTCGAAAGCTTACATGCTGTT 550  
DB 301 TTCACCTCAAGTTTGTGATCCGTAAATGAGCAAAATTCGAAAGCTTACATGCTGTT 360  
QY 551 GAGAGATTAATGAAACAATGCTTTGTAAGTCTTTTCAAGAGGAGGACCTTCGGAAGAG 610  
DB 361 GAGAGATTAATGAAACAATGCTTTGTAAGTCTTTTCAAGAGGAGGACCTTCGGAAGAG 420  
QY 611 GGCTGGCGCGGACAGCAACCTGCTGTCAACAGGAGCAACAGGAGATGAGAAGAGAGAG 670  
DB 421 GGCTGGCGCGGACAGCAACCTGCTGTCAACAGGAGCAACAGGAGATGAGAAGAGAGAG 480  
QY 671 GTGAGAGTGGCGGTCAAGTGGAGTGAACCTTCGAGCTTTCAGACCGGCTGCGAGAGTGC 730  
DB 481 GTGAGAGTGGCGGTCAAGTGGAGTGAACCTTCGAGCTTTCAGACCGGCTGCGAGAGTGC 540  
QY 731 TACCGGGTCAAGGCGCAGGACGTCAATGATGATG-AGAAGTCTGTTTGGCAGCCGAGAG 789





Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGTGCTGCGGCTCTGCGGCGCTAGTACACAGCCTGAGTGTGGCAACAAGAGAC 60
Db 1 AGTGCTGCGGCTCTGCGGCGCTAGTACACAGCCTGAGTGTGGCAACAAGAGAC 60
QY 61 CCTCTCATGTTTAAAGGACCTCTGCGGCTCAGAGAGCTGGCGGCGGCTCTGCGGAGAC 120
Db 61 CCTCTCATGTTTAAAGGACCTCTCTGCGGCTCAGAGAGCTGGCGGCGGCTCTGCGGAGAC 120
QY 121 TCCGCGCATCGCGGCGGCGCAATGCTCGGCGTGGCTGCGCAGTGTCTGCTGCTCCC 180
Db 121 TCCGCGCATCGCGGCGGCGCAATGCTCGGCGTGGCTGCGCAGTGTCTGCTGCTCCC 180
QY 181 TGGTGTCTGGGTCAAAAGTGTCTGGGTCTTGGGTCTTGGATTTGCGCGGCGGTTCAACGT 240
Db 181 TGGTGTCTGGGTCAAAAGTGTCTGGGTCTTGGGTCTTGGATTTGCGCGGCGGTTCAACGT 240
QY 241 AGCTGTGCGGCTCTCTGCGGTGAGTCCGTCCGCGCGCGGCTGCGCGGAGACCGGCTAGGC 300
Db 241 AGCTGTGCGGCTCTCTCTGCGGTGAGTCCGTCCGCGCGCGGCTGCGCGGAGACCGGCTAGGC 300
QY 301 TGCCTGGGAGTCCGCGGCGGCGCAATTCGCGGCTGCAATTCAGGAGATCCGCGATGCA 360
Db 301 TGCCTGGGAGTCCGCGGCGGCGCAATTCGCGGCTGCAATTCAGGAGATCCGCGATGCA 360
QY 361 CCGCGGCGGCGGCGGCGGCGCTTACCGAGCGGCTGCAAGCTGTGGGAGAGAGTCCGCGGAC 420
Db 361 CCGCGGCGGCGGCGGCGGCGCTTACCGAGCGGCTGCAAGCTGTGGGAGAGAGTCCGCGGAC 420
QY 421 GGGTCCCTGAGAGATCCGAGTGCCTACGAGCGCAAGATCTCAGCTTTATAGGTGTGACTA 480
Db 421 GGGTCCCTGAGAGATCCGAGTGCCTACGAGCGCAAGATCTCAGCTTTATAGGTGTGACTA 480
QY 481 CACATGTGACTTCACTCAGTTTGTGATCGTAAATGGAACAATTCGAAGCTACTTCA 540
Db 481 CACATGTGACTTCACTCAGTTTGTGATCGTAAATGGAACAATTCGAAGCTACTTCA 540
QY 541 CAGTGTCTGTGAAGATTAAATGAACAATTCGTGTAAGCTCTTTGCGAGAGAGACCC 600
Db 541 CAGTGTCTGTGAAGATTAAATGAACAATTCGTGTAAGCTCTTTGCGAGAGAGACCC 600
QY 601 TCGGAAGCAGAGGCTGCGGCGGCGAGAGCACTGTCTGTCAACAAGGACCAAGAGCAT 660
Db 601 TCGGAAGCAGAGGCTGCGGCGGCGAGAGCACTGTCTGTCAACAAGGACCAAGAGCAT 660
QY 661 GAAGACCCCGTGGAGTGGCCGTCACTGGGATGACAGACCTTCGCGCTTCAACAACGCTG 720
Db 661 GAAGACCCCGTGGAGTGGCCGTCACTGGGATGACAGACCTTCGCGCTTCAACAACGCTG 720
QY 721 CCGAGGTGCTTACCGGCTCAAGGCGAGGATGCAATGATGATGAGACTGTGTTGGCAG 780
Db 721 CCGAGGTGCTTACCGGCTCAAGGCGAGGATGCAATGATGATGAGACTGTGTTGGCAG 780
QY 781 CCGAGCAGAGCCTGCGGCTTACCCACCGAGCTTGCATCCGCTCTGGGTGAGAGAGCTTA 840
Db 781 CCGAGCAGAGCCTGCGGCTTACCCACCGAGCTTGCATCCGCTCTGGGTGAGAGAGCTTA 840
QY 841 CAGAACCAAGGCGTGGGCAAGAGGATGAAAGCTTGGGCGCAAAAGGAGCTGTGA 900
Db 841 CAGAACCAAGGCGTGGGCAAGAGGATGAAAGCTTGGGCGCAAAAGGAGCTGTGA 900
QY 901 GACCAACCCCTCAAGGCGGCGAGCAACCCCACTTCAACAAGGAGAGAAACAATTCAG 960
Db 901 GACCAACCCCTCAAGGCGGCGAGCAACCCCACTTCAACAAGGAGAGAAACAATTCAG 960
QY 961 ACCCATCAGACACCCGCTTACTGTGATGATGCTGTGGCTCCGATCTGAAG 1020
Db 961 ACCCATCAGACACCCGCTTACTGTGATGATGCTGTGGCTCCGATCTGAAG 1020
QY 1021 CGCCAGCTTGGGCGCCCGCGGATGGCGAAGGAGGAGATGCGCAAGCTCCGTCTCTTT 1080
Db 1021 CGCCAGCTTGGGCGCCCGCGGATGGCGAAGGAGGAGATGCGCAAGCTCCGTCTCTTT 1080
```

```
Db 1021 CGCCAGCTTGGGCGCCCGCGGATGGCGAAGGAGGAGATGCGCAAGCTCCGTCTCTTT 1080
QY 1081 GTGAGGCCAACCACTACCCCGAGGGGTAGCCACTGCGCCCGGCGCAAGGAGGCAACAT 1140
Db 1081 GTGAGGCCAACCACTACCCCGAGGGGTAGCCACTGCGCCCGGCGCAAGGAGGCAACAT 1140
QY 1141 GCGAGCCATTCACCCAGCTGTGCTCTCAAGACAGAGCGGCGGCGAGGAGACTCCCA 1200
Db 1141 GCGAGCCATTCACCCAGCTGTGCTCTCAAGACAGAGCGGCGGCGAGGAGACTCCCA 1200
QY 1201 GAAGTTATCTATGAGGTGAGTTTACACTTTCACGCCCCCTGAAGCGGAGATTTCCATT 1260
Db 1201 GAAGTTATCTATGAGGTGAGTTTACACTTTCACGCCCCCTGAAGCGGAGATTTCCATT 1260
QY 1261 CCTACCCACCTGAATGTCCCGAGCACTGAGTATCCAGGCAACAGAGGCCGCCCAACAA 1320
Db 1261 CCTACCCACCTGAATGTCCCGAGCACTGAGTATCCAGGCAACAGAGGCCGCCCAACAA 1320
QY 1321 TGGGCTCAGAGATCTCAGGCTTCCAGCTCAGGAGTGAACCTTCCGAGCCCTGTGTAC 1380
Db 1321 TGGGCTCAGAGATCTCAGGCTTCCAGCTCAGGAGTGAACCTTCCGAGCCCTGTGTAC 1380
QY 1381 TTCCAGGCTCGCTCAGTTAGCACTTTCAGTTCATTCACCCCAAGAGTGGGCGCAC 1440
Db 1381 TTCCAGGCTCGCTCAGTTAGCACTTTCAGTTCATTCACCCCAAGAGTGGGCGCAC 1440
QY 1441 CCAAGAACCAAGAGCCCTTGAAGATATCTTTCATTCAGGAGTGGCTATGGGCGCAC 1500
Db 1441 CCAAGAACCAAGAGCCCTTGAAGATATCTTTCATTCAGGAGTGGCTATGGGCGCAC 1500
QY 1501 GCGGACAGGATATGCGCCCTTGGCCAGGAGTGAAGAGCACTTATCACCAGGAAACCCAG 1560
Db 1501 GCGGACAGGATATGCGCCCTTGGCCAGGAGTGAAGAGCACTTATCACCAGGAAACCCAG 1560
QY 1561 TATTTAAAGAACCCCTGTGGGCGAGACAGACATACAGGAGGAGGAGGAGCTCTCTTT 1620
Db 1561 TATTTAAAGAACCCCTGTGGGCGAGACAGACATACAGGAGGAGGAGGAGCTCTCTTT 1620
QY 1621 ATCTGACATCTCTATGATGATTTCTTCCGATTTCCGATTTGGGAGGCGCAC 1680
Db 1621 ATCTGACATCTCTATGATGATTTCTTCCGATTTCCGATTTGGGAGGCGCAC 1680
QY 1681 CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAAATTAGTCCAAACGAG 1740
Db 1681 CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATATCTCAAAATTAGTCCAAACGAG 1740
QY 1741 GGCCTGGCACCTTCCGACATCTCAATTTGCTTGTGCGCAAGTCCGAATTAAGCGGCTGAT 1800
Db 1741 GGCCTGGCACCTTCCGACATCTCAATTTGCTTGTGCGCAAGTCCGAATTAAGCGGCTGAT 1800
QY 1801 TGGCAACTGTG 1811
Db 1801 TGGCAACTGTG 1811
```

RESULT 2  
AAH16354  
ID AAH16354 standard; cDNA; 1775 BP.

XX AAH16354;  
AC  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:15283.  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX



QY 1631 TCTTAGTCATCTTCTGCTTTTCTGCCGATTCGGGATTTGGGGGACCTCTAAGATG 1690  
DB 1441 TCTTAGTCATCTTCTGCTTTTCTGCCGATTCGGGATTTGGGGGACCTCTAAGATG 1500  
QY 1691 CCTCTCCAGCCCTGTCTCAACCATCTCCAAATTAATGAGCCACCCAGGGGCTTGAC 1750  
DB 1501 CCTCTCCAGCCCTGTCTCAACCATCTCCAAATTAATGAGCCACCCAGGGGCTTGAC 1560  
QY 1751 CTCCCATATCATCATCTGTCTGTGCGCAAGTGGAATTAACGGCGGTGATTCGCAACCTG 1810  
DB 1561 CTCCCATATCATCATCTGTCTGTGCGCAAGTGGAATTAACGGCGGTGATTCGCAACCTG 1620  
QY 1811 G 1811  
DB 1621 G 1621

RESULT 3  
ABA08585  
ID ABA08585 standard; cDNA; 1826 BP.  
XX ABA08585;  
AC  
XX  
DT 11-JAN-2002 (first entry)  
XX  
XX  
DE Human type II procollagen homologue-encoding cDNA, SEQ ID NO:361.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
XX haematopoietic regulation; tissue growth; immunomodulator; activin;  
XX inhibitor; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
XX chronic inflammatory condition; proliferative retinopathy;  
XX atherosclerosis; coronary heart disease; arterial ischaemia;  
XX bone disorder; osteoporosis; vascular growth disorder;  
XX tissue regeneration; wound healing; infection; immune disorder;  
XX cell culture; drug screening; gene therapy; antiinflammatory;  
XX antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
XX cytostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;  
XX antifungal; vulnery; antidiabetic; ss.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX  
XX PN MO200157188-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 05-FEB-2001; 2001MO-US003800.  
XX  
XX PR 03-FEB-2000; 2000US-00496914.  
XX  
XX PR 27-APR-2000; 2000US-00560875.  
XX  
XX PA (HYSE-) HYSEQ INC.  
XX  
XX PI Tang YT, Liu C, Drmanac RT;  
XX  
XX DR WPI: 2001-457740/49.  
XX  
XX DR P-PDB; ABB11341.  
XX  
XX PT Human proteins and DNA encoding sequences useful for preventing, treating  
XX or ameliorating a medical condition in a mammalian subject e.g. arthritis  
XX and cancer.  
XX  
XX PS Claim 1; Page 460; 1963pp; English.  
XX  
XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
XX CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
XX CC invention also relates to vectors and recombinant host cells comprising a  
XX CC nucleotide of the invention, methods of producing the novel polypeptides,  
XX CC antibodies against the polypeptides, methods of detecting the nucleotides,  
XX CC or polypeptides in a sample, and methods of identifying compounds which  
XX CC bind to polypeptides of the invention. Although novel, many of the

CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoietic regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a cDNA encoding a  
CC novel human polypeptide of the invention  
XX

SO Sequence 1826 BP; 382 A; 566 C; 515 G; 363 T; 0 U; 0 Other;

Query Match 88.9%; Score 1610.2; DB 4; Length 1826;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 197 AGTGTGGGTTCTGGGTTTCTGATTCGGGGGCGTTACACGTACCTGTGGGCTCC 256  
DB 10 ATTCTGGGTTCTGGGTTTCTGATTCGGGGGCGTTACACGTACCTGTGGGCTCC 69  
QY 257 TCGGGTGAATCCGTCCGCGCGCGGTGCCCGGAGCGGCTGAGCTCCGGGGTCCGGGG 316  
DB 70 TCGGGTGAATCCGTCCGCGCGCGGTGCCCGGAGCGGCTGAGCTCCGGGGTCCGGGG 129  
QY 317 CCCGAGCATTCGGGGCTGCAATTGACGGGGATCCCGGATGACCGCGGCCCGCGGC 376  
DB 130 CCCGAGCATTCGGGGCTGCAATTGACGGGGATCCCGGATGACCGCGGCCCGCGGC 189  
QY 377 CCTCACCGAGGGGTCCAGACTGTGGGAAAGAGTGGCGGGAGCGGGTCCCTGAGGATCC 436  
DB 190 CCTCACCGAGGGGTCCAGACTGTGGGAAAGAGTGGCGGGAGCGGGTCCCTGAGGATCC 249  
QY 437 CGATGCCCTACGAGCCAAAGTGTCAAGTTATAGTGTGACCTACCATGTGACCTTACCC 496  
DB 250 CGATGCCCTACGAGCCAAAGTGTCAAGTTATAGTGTGACCTACCATGTGACCTTACCC 309  
QY 497 TCAATTTTGTGATCCGTAATAATGACAAATTTGAAGCTACTTCAAGTGTGAGAGG 556  
DB 310 TCAATTTTGTGATCCGTAATAATGACAAATTTGAAGCTACTTCAAGTGTGAGAGG 369  
QY 557 ATTAAATGAACAATCTTTGTAAGGCTTTTGGAGGAGGAGCGGTGGAAGCGGGCTTG 616  
DB 370 ATTAAATGAACAATCTTTGTAAGGCTTTTGGAGGAGGAGCGGTGGAAGCGGGCTTG 429  
QY 617 GCGGAGAGCAACCTGTGTTCACGAGGAGCAACGAGGATGAAGACCCCGGTGAG 676  
DB 430 GCGGAGAGCAACCTGTGTTCACGAGGAGCAACGAGGATGAAGACCCCGGTGAG 489  
QY 677 CTGGCGGTGAGTGGATGACAGCCTTGGGCTTTCAACACCGCTGCCAGAGTGGCTACCG 736  
DB 490 CTGGCGGTGAGTGGATGACAGCCTTGGGCTTTCAACACCGCTGCCAGAGTGGCTACCG 549



```

QY 737 GTCAAGCCAGACGTCATATGTGATGAGACTCTGTTGGAGCCCAAGCAACCCG 796
DB 550 GTCAAGCCAGACGTCATATGTGATGAGACTCTGTTGGAGCCCAAGCAACCCG 609
QY 797 CTTACCCCAACCGGACTTGGATCCGCCCTGGTGAAGAGGCTTAACGAACCAAGAGCCGTG 856
DB 610 CTTACCCCAACCGGACTTGGATCCGCCCTGGTGAAGAGGCTTAACGAACCAAGAGCCGTG 669
QY 857 GGCAGAGGAGGATGAGAGGCTTGGGGGGAAGGGGAGGAGTGTGAGACCCCTCCCAAG 916
DB 670 GGCAGAGGAGGATGAGAGGCTTGGGGGGAAGGGGAGGAGTGTGAGACCCCTCCCAAG 729
QY 917 GGCAGACACCCCAACCTCAACCAAGAGAGAAACAAATAGACACCCATCAAGCAAC 976
DB 730 GGCAGACACCCCAACCTCAACCAAGAGAGAGAAACAAATAGACACCCATCAAGCAAC 789
QY 977 CCGTCTTACTGTGATGAGTCCGCTGTTGGCTCCGATCTGAAGGCGCCAGCTTGGGGCC 1036
DB 790 CCGTCTTACTGTGATGAGTCCGCTGTTGGCTCCGATCTGAAGGCGCCAGCTTGGGGCC 849
QY 1037 CCGGAGATGGGGAAGGGGGAATGCGCAAGCTCCGCTCTTGTGAGACCCCAACCT 1096
DB 850 CCGGAGATGGGGAAGGGGGAATGCGCAAGCTCCGCTCTTGTGAGACCCCAACCT 909
QY 1097 ACCCCAGGGGAGGACACTCGCCCGCCCAAGGAGGACCACTGAGCCATTACCCA 1156
DB 910 ACCCCAGGGGAGGACACTCGCCCGCCCAAGGAGGACCACTGAGCCATTACCCA 969
QY 1157 GCTGTCCTCTCAAGACAGAGCGGGGCGCAGGAGACTCCAGAGTTATCTATGAGT 1216
DB 970 GCTGTCCTCTCAAGACAGAGCGGGGCGCAGGAGACTCCAGAGTTATCTATGAGT 1029
QY 1217 GGGTTACACTCTTCAACGCCCCCTGAAGGGGGAATTTCCATCTCCCTCAACCTGAT 1276
DB 1030 GGGTTACACTCTTCAACGCCCCCTGAAGGGGGAATTTCCATCTCCCTCAACCTGAT 1089
QY 1277 GTCCCAAGCACTGTGATCAGACCAAGTGGCCCCCAACAAATGGGCTCAGGATCTC 1336
DB 1090 GTCCCAAGCACTGTGATCAGACCAAGTGGCCCCCAACAAATGGGCTCAGGATCTC 1149
QY 1337 AGGCTTTCACGTCAGGGGATGACTTCCGAGACCCCTGTGATCTTCCAGGGCTGCTCA 1396
DB 1150 AGGCTTTCACGTCAGGGGATGACTTCCGAGACCCCTGTGATCTTCCAGGGCTGCTCA 1209
QY 1397 GTTAGCATTTCAATGTCCTTAACCCCAAGCAAGTGGGGCCACCAAGAAACCAAGCCC 1456
DB 1210 GTTAGCATTTCAATGTCCTTAACCCCAAGCAAGTGGGGCCACCAAGAAACCAAGCCC 1269
QY 1457 CTTGGAATGATGATCTTTTCAATGAGGTTGCTATGGGGCCACAGGATATGAGCC 1516
DB 1270 CTTGGAATGATGATCTTTTCAATGAGGTTGCTATGGGGCCACAGGATATGAGCC 1329
QY 1517 CTTTCCAGGGATGAGAGCAATTCATCAACCAAGGAAACCCAGATTAAGAAAGCCCT 1576
DB 1330 CTTTCCAGGGATGAGAGCAATTCATCAACCAAGGAAACCCAGATTAAGAAAGCCCT 1389
QY 1577 GTGGGGGAGACAGACATAGCAAGGGGTGGGCACTGCTCTTATCTGACATCTCTA 1636
DB 1390 GTGGGGGAGACAGACATAGCAAGGGGTGGGCACTGCTCTTATCTGACATCTCTA 1449
QY 1637 GTGCAATCTTGCTTTTCTCCGATGCGGATTTGGGGGCAACCTTAAGATGCTCTC 1696
DB 1450 GTGCAATCTTGCTTTTCTCCGATGCGGATTTGGGGGCAACCTTAAGATGCTCTC 1509
QY 1697 TCCAGCCCTGTCTCAACCATATCTCCAAATAGTGCAACCAAGGGGCTGAGCACTCCCA 1756
DB 1510 TCCAGCCCTGTCTCAACCATATCTCCAAATAGTGCAACCAAGGGGCTGAGCACTCCCA 1569
QY 1757 CATATCATATGCTGTGCTGCAAGTGGCAATAAACGGCGTATGTCACCACTGG 1811
DB 1570 CATATCATATGCTGTGCTGCAAGTGGCAATAAACGGGTGATGTCACCACTGG 1624

```

```

RESULT 4
AAH16384
ID AAH16384 standard; cDNA; 1614 BP.
XX
AC AAH16384;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15334.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isegai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 15334; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH1629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 1614 BP; 317 A; 514 C; 487 G; 296 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 1434.8; DB 4; Length 1614;
Best local similarity 90.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
QY 25 AGTACACACGACCTGAGTGTGAGGACAGAGAGACCTCTCATGTTTAAAGACCTCT 84
DB 1 AGTACACACGACCTGAGTGTGAGGACAGAGAGACCTCTCATGTTTAAAGACCTCT 60

```

```

QY 85 GGGGCTCAGAGCGTGGCGCCCGCCCTGGGCGGACTCCCCCATCCGCGGCGCGAATG 144
Db 61 GGGGCTCAGAGCGTGGCGCCCGCCCTGGGCGGACTCCCCCATCCGCGGCGCGAATG 120
QY 145 GTCCGGGTCGGCTCCGAGTGTCTGTGCTGCTCCCTGGTGTGCTGGTGAAGTGTG 204
Db 121 GTCCGGGTCGGCTCCGAGTGTCTGTGCTGCTCCCTGGTGTGCTGGTGAAGTGTG 180
QY 205 GTTCTGGGTTTCTGGATTTCGGGCGGTTCAAGTGAAGCTGTGCGGCTCTCCGTGTA 264
Db 181 GTTCTGGGTTTCTGGATTTCGGGCGGTTCAAGTGAAGCTGTGCGGCTCTCCGTGTA 240
QY 265 GTCCGTCGCGCGCGTGGCCCGGAGCGGCTAGGCTGCGCGGAGTCCGCGGCGCCAGG 324
Db 241 GTCCGTCGCGCGCGTGGCCCGGAGCGGCTAGGCTGCGCGGAGTCCGCGGCGCCAGG 300
QY 325 ATTCCGGGCTGCAATGAGCGGAGATCCCGATGACCGCGCGCCCGCGCTCAACG 384
Db 301 ATTCCGGGCTGCAATGAGCGGAGATCCCGATGACCGCGCGCCCGCGCTCAACG 360
QY 385 ACCGGTCCAGACCTGGTGGGAAGGTTGGGCGGAGCGGCTCCGAGGATCCCGATGCT 444
Db 361 ACCGGTCCAGACCTGGTGGGAAGGTTGGGCGGAGCGGCTCCGAGGATCCCGATGCT 420
QY 445 ACAGCCAGATGCTCAGCTTTATAGTGTGACTTACATGTAATGACTTCACTCAATTT 504
Db 421 ACAGCCAGATGCTCAGCTTTATAGTGTGACTTACATGTAATGACTTCACTCAATTT 480
QY 505 GTGATCCGTAAATGCAAAATTCGAGCTTCAAGTGTGTTGAAGATTAATG 564
Db 481 GTGATCCGTAAATGCAAAATTCGAGCTTCAAGTGTGTTGAAGATTAATG 540
QY 565 AAACAATGCTTGTAAAGCTTTTGCAGAGGAGGCTCCGAGAGGAGGCTCCGCGGAG 624
Db 541 AAACAATGCTTGTAAAGCTTTTGCAGAGGAGGCTCCGAGAGGAGGCTCCGCGGAG 600
QY 625 AGCAACCTGCTGTCAACAGGAGCAACAGCAGCAGTGAAGAGCCCTGTGAGTGTGCT 684
Db 601 AGCAACCTGCTGTCAACAGGAGCAACAGCAGCAGTGAAGAGCCCTGTGAGTGTGCT 660
QY 685 CAGTGGGATGCAACCTCCGCTTCAACAGCAGCAGTGTGAGTGTGAGTGTGAGTGTG 744
Db 661 CAGTGGGATGCAACCTCCGCTTCAACAGCAGCAGTGTGAGTGTGAGTGTGAGTGTG 720
QY 745 CAGGAGCTCATGTGATGAGTCTGTTTGGGAGGCGGAGGAGGCGGCTCAACCC 804
Db 721 CAGGAGCTCATGTGATGAGTCTGTTTGGGAGGCGGAGGAGGCGGCTCAACCC 780
QY 805 ACCGGAATTGATCCGCTGTGGTGAAGAGGCTTACAGAACAGAGCGTGGGCAAGA 864
Db 781 ACCGGAATTGATCCGCTGTGGTGAAGAGGCTTACAGAACAGAGCGTGGGCAAGA 840
QY 865 GGCATGGAAGGCTTGGGCGCAAGGAGGAGTGTGAGACACCCCTTCAAGGGCGAGAC 924
Db 841 GGCATGGAAGGCTTGGGCGCAAGGAGGAGTGTGAGACACCCCTTCAAGGGCGAGAC 900
QY 925 CCCCACTTCAACAGAGAGAGCAATACAGACCATGACGACACCCCGTCTTA 984
Db 901 CCCCACTTCAACAGAGAGAGCAATACAGACCATGACGACACCCCGTCTTA 960
QY 985 CTGTGATGATGCTGTTTGGCTCCGATGTGAAGGCGCAGCTTGGGCGCCCGCGAT 1044
Db 961 CTGTGATGATGCTGTTTGGCTCCGATGTGAAGGCGCAGCTTGGGCGCCCGCGAT 1020
QY 1045 GGGGAGGGGAGTCCGCAAGTCTCGTCTTGTGAGCGCACCACTTACCCGAG 1104
Db 1021 GGGGAGGGGAGTCCGCAAGTCTCGTCTTGTGAGCGCACCACTTACCCGAG 1080
QY 1105 GGGGAGGCACTGGCCCGCGCCGAGGAGGAGCAGCAGTTCACCACTGCTTC 1164
Db 1081 GGGGAGGCACTGGCCCGCGCCGAGGAGG----- 1109
QY 1165 CTCGAAGACAGAGCCGGGCGCAGCGGAGACTCCAGAGATTATATGAGTGTGCA 1224

```

```

Db 1110 ----- 1109
QY 1225 CTCCTACGCCCCCTGAAGCGGGAATTTCCTCCATTCCTCAGCCAGTATGCCAG 1284
Db 1110 ----- 1109
QY 1285 CACTGGTATCCAGCCAGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1344
Db 1110 -----CAGCAGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1154
QY 1345 CAGTCAAGGAGTACCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1404
Db 1155 CAGTCAAGGAGTACCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1214
QY 1405 TTGAGTGCATCTAACCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1464
Db 1215 TTGAGTGCATCTAACCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1274
QY 1465 ATGATCTCTTTTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1524
Db 1275 ATGATCTCTTTTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1334
QY 1525 GGGTAGAGGAGATTCATCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1584
Db 1335 GGGTAGAGGAGATTCATCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1394
QY 1585 AGACAGACATAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1644
Db 1395 AGACAGACATAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1454
QY 1645 TTGCTCTTTTCTCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1704
Db 1455 TTGCTCTTTTCTCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1514
QY 1705 TGTCTAACCATCTCAATTTAGTCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAG 1764
Db 1515 TGTCTAACCATCTCAATTTAGTCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAG 1574
QY 1765 ATTGTCTGTGCGCAAGTGTGAATTAAGCGGCTGATTC 1804
Db 1575 ATTGTCTGTGCGCAAGTGTGAATTAAGCGGCTGATTC 1614

RESULT 5
AD057486
ID AD057486 standard; cDNA, 709 BP.
XX
AC AD057486;
XX
DT 29-JUL-2004 (first entry)
XX
DE DNA encoding hairless protein interaction partner protein #45.
XX
KW sb; gene; human; hairless protein; Hrc; ubiquitons receptor UR; MAP1A;
KW KIAA0930 protein; monocyes antigen CD14; sphingolipid activator protein;
KW beta-synuclein; C11 protein; vesicle-associated membrane protein 2;
KW aldolase A; Cgl-106 protein; hypothalamus protein HSMN1; alpha enolase;
KW POM-2P3; quinone oxidoreductase; pumilio 1; VPS41; KIAA0614 protein;
KW splicing factor Ccl.4; ubiquitin; beta-mannosidase; hair growth.
XX
OS Homo sapiens.
XX
PN US2004086945-A1.
XX
PD 06-MAY-2004.
XX
PF 02-JUN-2003; 2003US-00452858.
XX
PR 03-JUN-2002; 2002US-0385414P.
XX
PA (PROC ) PROCTER & GAMBLE CO.
XX

```

P1 Sreekrishna K, Gerwe GS, Toerner DR;  
XX WPI: 2004-430095/40.  
DR P-PSDB; AD057487.  
XX  
PT New composition comprising mouse truncated hairless protein-human  
PT interacting partner protein or nucleic acid complexes, useful for  
PT screening test compounds that inhibit or enhance hair growth.  
XX  
PS Claim 6; SEQ ID NO 78; 60pp; English.  
XX  
CC The invention relates to a composition comprising a mouse truncated  
CC hairless (Hr) protein-human interacting partner protein or nucleic acid  
CC complex. The human interacting partner protein comprises a molecule  
CC selected from ubiquitous receptor UR, MAP1A, KIAA0930 protein, monocytes  
CC antigen CD14, sphingolipid activator protein, beta-synuclein, C11  
CC protein, vesicle-associated membrane protein 2, aldolase A, CGI-106  
CC protein, hypohalimus protein HSNMP1, alpha enolase, POM-2P3, quinone  
CC oxidoreductase, pumilio 1, VPS41, KIAA0614 protein, splicing factor  
CC C1.4, ubiquitin, beta-mannosidase. Also described are: methods of  
CC assaying a test compound for agonist or antagonist activity for the above  
CC composition; and method of inhibiting or increasing hair growth on a  
CC surface in a subject. The present sequence represents DNA encoding a  
CC hairless protein human interacting partner used in the composition.  
XX  
SQ Sequence 709 BP; 146 A; 247 C; 203 G; 110 T; 0 U; 3 Other;

Query Match 38.0%; Score 688.2; DB 12; Length 709;  
Best Local Similarity 99.1%; Pred. No. 8.7e-176;  
Matches 690; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 682 CGTCAGTGGATGACGACCTCGGCTTACAGACCGCTGGAGGTGCTACCGGGCTAA 741  
DB 14 CGTCAGGAGGATGACGACCTCGGCTTACAGACCGCTGGAGGTGCTACCGGGCTAA 73  
QY 742 GGGCAGGAGCTCATATGTGATGATGAGACTCTGTTGGAGCCGAGGACACCGGCTTAC 801  
DB 74 GGGCAGGAGCTCATATGTGATGATGAGACTCTGTTGGAGCCGAGGACACCGGCTTAC 133  
QY 802 CCCACCGGACTTCCATCGGCTTGGTGGAGAGGCTTAAAGAACCAAGGCGTGGGCA 861  
DB 134 CCCACCGGACTTCCATCGGCTTGGTGGAGAGGCTTAAAGAACCAAGGCGTGGGCA 193  
QY 862 GGAGGAGCTGAAGGCTTGGGGGCAAGGGAGCTGTAGACCCCTTCAAGGGGGAG 921  
DB 194 GGAGGAGCTGAAGGCTTGGGGGCAAGGGAGCTGTAGACCCCTTCAAGGGGGAG 253  
QY 922 CACCCGCCCTTCAACCAAGAGAGAAATACAGACCCATCAGCCACACCCGCTC 981  
DB 254 CACCCGCCCTTCAACCAAGAGAGAAATACAGACCCATCAGCCACACCCGCTC 313  
QY 982 TTACTGTGATGATGCTGCTGTTGGCTCCGATCTGAGAGGCGGCTTGGGGGCGCG 1041  
DB 314 TTACTGTGATGATGCTGCTGTTGGCTCCGATCTGAGAGGCGGCTTGGGGGCGCG 373  
QY 1042 GATGCGGAAGGGGGATGCGGAAGCTCGGCTCTGTTGAGCGGACCACTTACCC 1101  
DB 374 GATGCGGAAGGGGGATGCGGAAGCTCGGCTCTGTTGAGCGGACCACTTACCC 433  
QY 1102 CAGGGGTAGCCACTCGCCCGCCCGGAGGAGGACCACTGCGAGCTTCAACCCAGCTGG 1161  
DB 434 CAGGGGTAGCCACTCGCCCGCCCGGAGGAGGACCACTGCGAGCTTCAACCCAGCTGG 493  
QY 1162 TCCCTCCAGAGCAAGCGGGGGCCAGCGGCAAGCTCCGAAAGTTATGTATGGTGGTT 1221  
DB 494 TCCCTCCAGAGCAAGCGGGGGCCAGCGGCAAGCTCCGAAAGTTATGTATGGTGGTT 553  
QY 1222 ACATCTTCAAGCGCCCGCTGAAGCGGGGACTTCCCATTTCCCTCAACCCAGCTGAAGTCC 1281  
DB 554 ACATCTTCAAGCGCCCGCTGAAGCGGGGACTTCCCATTTCCCTCAACCCAGCTGAAGTCC 613  
QY 1282 CAGCACTGTGATCAGGACCAAGTGGCCCGGACCAAGTGGGCTCAGATCTCAGGCGC 1341

DB 614 CAGCACTGTGATCAGGACCAAGCGGCGCCCGGACCAAGTGGGCTCAGATCTCAGGCGC 673  
QY 1342 TTCCACGTCGAGGGGAGACTTCCGAGAGCCCTTGCT 1377  
DB 674 TTTCACGTCGAGGGGAGACTTCCGAGAGCCCTTGCT 709

RESULT 6  
AAH07100  
ID AAH07100 standard; cDNA; 784 BP.  
XX  
AC AAH07100;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:3935.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
FN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PE 28-JUN-2000; 2000EP-00116126.  
XX  
PR 29-JUN-1999; 99JP-00248036.  
XX  
PR 27-AUG-1999; 99JP-00300253.  
XX  
PR 11-JAN-2000; 2000JP-00118776.  
XX  
PR 02-MAY-2000; 2000JP-00183767.  
XX  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
DR WPI: 2001-318749/34.  
XX  
PT primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.  
XX  
PS Claim 1; SEQ ID NO 3935; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC nucleotide strand of a polynucleotide which comprises one of the 5602  
CC complementary sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 784 BP; 174 A; 224 C; 251 G; 130 T; 0 U; 5 Other;

Query Match 36.5%; Score 661.4; DB 4; Length 784;









497 CCCCCCANAAATGGGCTGAGATNTCAGGCTTCCACGTCAGGGGTGACTTCCGGA 438  
1368 GCGCCCTGTGACTTCCAGGGGTGCTGCTAGTTAGATTGAGTGCATCTACCCACGAC 1427  
437 GCGCCCTGTGACTTCCAGGGGTGCTGCTAGTTAGATTGAGTGCATCTACCCACGAC 378  
1428 GAGGTGGGGCCACCCAGAAACCAAGCCCTTGAATAGATCTTTTCATCAGGGTTG 1487  
377 GAGGTGGGGCCACCCAGAAACCAAGCCCTTGAATAGATCTTTTCATCAGGGTTG 318  
1488 CCTATGGGGCCACCGCGACAGGTATGCCCCCTTGCAGGGTGAAGACATTCATCACCC 1547  
317 CCTATGGGGCCACCGCGACAGGTATGCCCCCTTGCAGGGTGAAGACATTCATCACCC 258  
1548 AGGGAACCCCGAGTATTAAAGAGCCCTGTGGGGCGACAGACATGACAGGGGTGAGC 1607  
257 AGGGAACCCCGAGTATTAAAGAGCCCTGTGGGGCGACAGACATGACAGGGGTGAGC 198  
1608 AGTGCCTCCCTTATCTGACATATCTAGTGGATTCTTCTCCCGATTGGGG 1667  
197 AGTGCCTCCCTTATCTGACATATCTAGTGGATTCTTCTCCCGATTGGGG 138  
1668 ATTGGGGGCGACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATCTCAAAATTA 1727  
137 ATTGGGGGCGACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCATCTCAAAATTA 78  
1728 GTGCCAACCCAGGGGCTGTGGACCTTCCACATCATCTGTTCTTGGCCAAAGTGGAA 1787  
77 GTGCCAACCCAGGGGCTGTGGACCTTCCACATCATCTGTTCTTGGCCAAAGTGGAA 18  
1788 TAAACGGCGGATGGCC 1804  
17 TAAACGGCGGATGGCC 1

## RESULT 11

AD116303/C  
ID AD116303 standard; DNA; 3673 BP.

AC AD116303;

DT 22-APR-2004 (first entry)

DE Human nucleic acid-associated protein (NAAP) coding sequence #38.

XX human; nucleic acid-associated protein; NAAP; autoimmune disorder;  
XX inflammatory disorder; AIDS; allergy; infection; metabolic disorder;  
XX obesity; reproductive disorder; infertility; neurological disorder;  
XX Parkinson's disease; Alzheimer's disease; cardiovascular disorder;  
XX myocardial infarction; hypertension; eye disorder;  
XX cell proliferative disease; cancer; ds; gene.

OS Homo sapiens.

XX MO2003094848-A2.

XX 20-NOV-2003.

XX 09-MAY-2003; 2003WO-US014450.

XX 10-MAY-2002; 2002US-0379843P.

XX 24-MAY-2002; 2002US-0383457P.

XX 31-MAY-2002; 2002US-0384699P.

XX 06-JUN-2002; 2002US-0387265P.

PA (INCY-) INCYTE CORP.

XX Kable AE, Elliott VS, Tran UK, Ramkumar J, Marquis JP, Chawla NK;  
PI Richardson TW, Bulloch SA, Khare R, Lee SY, Lai PG, Tang YT, Yue H;  
PI Swarnakar A, Beecha SD, Hafalia AJA, Chang H, Baughn MR, Borowsky ML;  
PI Gietzen KJ, He A, Forsythe J, Sprague WM, Blake JT, Warren BA;  
PI Mason PM, Ison CH, Lindquist BA, Wilson AD, Jin P;

XX MPI: 2004-011999/01.  
DR P-PSDB: AD116253.  
XX  
XX New human nucleic acid associated proteins and polynucleotides, useful  
PT for diagnosing, preventing or treating diseases or conditions associated  
PT with aberrant protein expression, e.g. cancer, AIDS, atherosclerosis or  
PT stroke.  
XX  
XX Claim 5; SEQ ID NO 88; 400bp; English.

XX The invention comprises the amino acid and coding sequences of human  
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
CC the invention are useful in diagnosing, preventing and treating  
CC diseases/conditions associated with altered expression of NAAP, such as:  
CC autoimmune/inflammatory disorders (e.g. AIDS and allergies), infections  
CC (e.g. bacterial and viral), metabolic disorders (e.g. obesity),  
CC reproductive disorders (e.g. infertility), neurological disorders (e.g.  
CC Parkinson's disease and Alzheimer's disease), cardiovascular disorders  
CC (e.g. myocardial infarction and hypertension), eye disorders, or cell  
CC proliferative diseases (e.g. cancer). The present DNA sequence encodes a  
CC human NAAP protein of the invention.

XX Sequence 3673 BP; 785 A; 1129 C; 1196 G; 563 T; 0 U; 0 Other;

Query Match 26.6%; Score 481.4; DB 12; Length 3673;

Best Local Similarity 99.6%; Pred. No. 1.7e-119; Matches 493; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 AGTGCCTGCGGCGCTTGGGCGGCTTGTACACACGACCTGAGTGAAGGACAGAGAC 60  
495 AGTGCCTGCGGCGCTTGGGCGGCTTGTACACACGACCTGAGTGAAGGACAGAGAC 436  
61 CCTTCCATGTTTAAAGGACCTCTGAGGCTTGAAGAGCGGCGGCGGCGGAC 120  
435 CCTTCCATGTTTAAAGGACCTCTGAGGCTTGAAGAGCGGCGGCGGCGGCGGAC 376  
121 TCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
375 TCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 316  
181 TGGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
315 TGGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 256  
241 AGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
255 AGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 196  
301 TGCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
195 TGCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 136  
361 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 419  
135 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 76  
420 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 479  
75 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 16  
480 ACACATGTGACTTCA 494  
15 ACACATGTGACTTCA 1

## RESULT 12

ACH14455  
ID ACH14455 standard; cDNA; 477 BP.

AC ACH14455;

DT 13-OCT-2003 (first entry)

XX DE Human adult brain cDNA #1667.  
 XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 XX KW genome mapping; biodiversity; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN US2003073623-A1.  
 XX PD 17-APR-2003.  
 XX PF 30-JUL-2001; 2001US-00918995.  
 XX PR 30-JUL-2001; 2001US-00918995.  
 XX PA (DRMA/) DRMANAC R T.  
 XX PA (LABA/) LABAT I.  
 XX PA (STAC/) STACHE-CRAIN B.  
 XX PA (DICK/) DICKSON M C.  
 XX PA (JONE/) JONES L W.  
 XX PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
 XX DR WPI; 2003-615964/58.  
 XX DR  
 XX PT New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.  
 XX PS Claim 1; SEQ ID NO 1667; 44pp; English.  
 XX XX  
 CC The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docid=20030073623  
 XX CC  
 SQ Sequence 477 BP; 86 A; 174 C; 125 G; 90 T; 0 U; 2 Other;

Query Match 24.1%; Score 436.6; DB 9; Length 477;  
 Best Local Similarity 99.1%; Pred. NO. 1e-107;  
 Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 986 TGGAGAGAGTCGCTGTTGGCTCCCATGGAAGGCGCCAGCTTCGGGGCCCGCGGAG 1045  
 DB 35 TGGAAATTCGTCGCTGTTGGCTCCCATGGAAGGCGCCAGCTTCGGGGCCCGCGGAG 94  
 QY 1046 GCGAAGGGGATCCGCAAGACTCCGTCCTTGTGAGAGCCACCACTTACCCGAG 1105  
 DB 95 GCGAAGGGGATCCGCAAGACTCCGTCCTTGTGAGAGCCACCACTTACCCGAG 154  
 QY 1106 GGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCACTTACCCAGTGTGCC 1165  
 DB 155 GGTAGCCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCACTTACCCAGTGTGCC 214  
 QY 1166 TCCAAAGACAGAGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1225  
 DB 215 TCCAAAGACAGAGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 274

QY 1226 TCTTCAAGCCCCCTGGAAGGGGAGACTTCCATTCCCTGACCCAGCTGAATGTCGCCAG 1285  
 DB 275 TCTTCAAGCCCCCTGGAAGGGGAGACTTCCATTCCCTGACCCAGCTGAATGTCGCCAG 334  
 QY 1286 ACTGGTCATTCAGACCAACAGTGGCCCCCAACAATAGGCTTGAAGTCTCAGGCTTCC 1345  
 DB 335 ACTGGTCATTCAGACCAACAGTGGCCCCCAACAATAGGCTTGAAGTCTCAGGCTTCC 394  
 QY 1346 ACCTCAGGGGTGACTTCCCGAGAGCCCCCTGGTGAATCTTCCAGGGCTGCTCAGTTACATT 1405  
 DB 395 ACCTCAGGGGTGACTTCCCGAGAGCCCCCTGGTGAATCTTCCAGGGCTGCTCAGTTACATT 454  
 QY 1406 TCAGTCCATCTACCCACGACG 1428  
 DB 455 TCAGTCCATCTACCCACGACG 477

RESULT 13  
 ACH3954  
 ID ACH43954 standard; cDNA; 484 BP.  
 AC ACH43954;  
 XX  
 DT 13-OCT-2003 (first entry)  
 XX  
 DE Human foetal brain cDNA #4679.  
 XX  
 KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 KW genome mapping; biodiversity; genetic disorder.  
 XX OS Homo sapiens.  
 XX OS  
 XX PN US2003073623-A1.  
 XX PD 17-APR-2003.  
 XX PF 30-JUL-2001; 2001US-00918995.  
 XX PR 30-JUL-2001; 2001US-00918995.  
 XX PA (DRMA/) DRMANAC R T.  
 XX PA (LABA/) LABAT I.  
 XX PA (STAC/) STACHE-CRAIN B.  
 XX PA (DICK/) DICKSON M C.  
 XX PA (JONE/) JONES L W.  
 XX PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
 XX DR WPI; 2003-615964/58.  
 XX DR  
 XX PT New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.  
 XX PS Claim 1; SEQ ID NO 31166; 44pp; English.  
 XX XX  
 CC The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at

```
CC seqdata.uspto.gov/sequence.html?docID=20030073623
XX
SQ Sequence 484 BP; 118 A; 129 C; 137 G; 85 T; 0 U; 15 Other;
Query Match 22.7%; Score 410.6; DB 9; Length 484;
Best Local Similarity 93.9%; Pred. No. 1.1e-100;
Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 426 CCTAGAGATCCCGATGCTTCAAGCCAGAGATGCTTCAAGCTTATAGGTGACCTTACAT 485
DB 39 CANNANANNNNNNNANANANACAGCCAGATGCTCACTTATAGGTGACCTTACTAT 98
QY 486 GTGACTTACCTCACTGATTTGTGATCCGTAAATGAGCAAAATTGCAAGCTACTTCAAGTG 545
DB 99 GTGACTTACCTCACTGATTTGTGATCCCAAAATGAGCAAAAGCCAGAGCTACTTCAAGAG 158
QY 546 CTGTTGAGAGATTAATGAAAACAATGCTTGAAGCTTTTTCAGAGAGGAGCCTCGGA 605
DB 159 CTGTTGAGAGATTAATGAAAACAATGCTTGAAGCTTTTTCAGAGAGGAGCCTCGGA 218
QY 606 AGCAGAGGCTGCGCCGCGAGAGACACCTGCTGACACAGGACCAAGGACATGAGA 665
DB 219 AGCAGAGGCTGCGCCGCGAGAGACACCTGCTGACACAGGACCAAGGACATGAGA 278
QY 666 CCCCCTGAGAGCTGAGCGCTCAAGTGGAGATGACAGCCCTCGGCTTACAGCAGCCTCGAG 725
DB 279 CCCCCTGAGAGCTGAGCGCTCAAGTGGAGATGACAGCCCTCGGCTTACAGCAGCCTCGAG 338
QY 726 GTGCTTACCGGGTCAAGGCCAGAGCGTCAATATGTAATGAGACTGTGTTGGAGCCGAG 785
DB 339 GTGCTTACCGGGTCAAGGCCAGAGCGTCAATATGTAATGAGACTGTGTTGGAGCCGAG 398
QY 786 CAGGACACCCGCTTACCCCAACCGGACTTGCATCCGCTTGGGTGAGAAAGGCTAACAGA 845
DB 399 CAGGACACCCGCTTACCCCAACCGGACTTGCATCCGCTTGGGTGAGAAAGGCTAACAGA 458
QY 846 CCAAGAGCGTGGGCAAGGAGGAGCATCG 871
DB 459 CCAAGAGCGTGGGCAAGGAGGAGCATCG 484

RESULT 14
ADP81362
ID ADP81362 standard; DNA; 572 BP.
AC ADP81362;
AC 26-FEB-2004 (first entry)
DT 26-FEB-2004 (first entry)
XX
XX Leukaemia-related DNA sequence #1918.
XX
XX Cytostatic; Gene therapy; leukaemia; ss.
XX
XX Unidentified.
XX
XX WO2003039443-A2.
XX
XX 15-MAY-2003.
XX
XX 04-NOV-2002; 2002MO-EP012303.
XX
XX 05-NOV-2001; 2001EP-00126244.
XX
XX 30-APR-2002; 2002EP-00009758.
XX
XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
XX (UYLU-) UNIV LUDWIG MAXIMILIANS.
XX
XX (HAPE/) HAFERLACH T.
XX
XX (SCHN/) SCHUCH C.
XX
XX (KERN/) KERN W.
XX
XX Haferlach T, Schoch C, Kern W, Kohlmann A, Schittger S, Dugas M,
XX
XX Eile R, Brose B, Mergenthaler S;
```

```
DR WPI; 2003-505037/47.
XX
XX Determining the subtype of leukemia cells and whether a patient sample
XX contains leukemia cells or other cells, useful for treating leukemia,
XX PT comprises determining the expression profile of a group of markers in a
XX patient sample.
XX
XX Disclosure; SEQ ID NO 1918; 2938bp; English.
XX
XX The present invention relates to a method (M1) for determining the
XX CC subtype of leukemia cells and whether a patient sample contains
XX CC leukemia cells. The method comprises determining the expression profile
XX of a group of markers in a patient sample. The method is useful for
XX CC determining the presence of leukemia cells, its types or subtypes, and
XX for the preparation of a medicament for treating leukaemia.
XX
SQ Sequence 572 BP; 113 A; 154 C; 129 G; 126 T; 0 U; 50 Other;
Query Match 20.4%; Score 369; DB 10; Length 572;
Best Local Similarity 94.9%; Pred. No. 2.3e-89;
Matches 369; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1397 GTTACATTTTCAGTCCATCTTACCCACAGAGGAGGAGGACCCAGAAACCAAGCCC 1456
DB 1 GTTACATTTTCAGTCCATCTTACCCACAGAGGAGGAGGACCCAGAAACCAAGCCC 60
QY 1457 CTTGGAATATGATCTCTTTTATCAGGGTTCCTTATGGGGCCACAGGATATGGCC 1516
DB 61 CTTGGAATATGATCTCTTTTATCAGGGTTCCTTATGGGGCCACAGGATATGGCC 120
QY 1517 CTTGCGAGGATGAGGAGACATTCATCACCCAGGAAACCCAGGATTAAGAAAGCCCT 1576
DB 121 CTTGCGAGGATGAGGAGACATTCATCACCCAGGAAACCCAGGATTAAGAAAGCCCT 180
QY 1577 GTGGGGGAGACAGACATAGCAGGGGTGGGAGTGCCTCCTTTATTCCTGACATCTTA 1636
DB 181 GTGGGGGAGACAGACATAGCAGGGGTGGGAGTGCCTCCTTTATTCCTGACATCTTA 240
QY 1637 GTCGATTTTGGCTTTTCTCCGATTCGGATTTGGGGGACACCTCTTAAGATGCTCTC 1696
DB 241 GTCGATTTTGGCTTTTCTCCGATTCGGATTTGGGGGACACCTCTTAAGATGCTCTC 300
QY 1697 TCCAGCCCTGTCTAACCATCTCCAAATTAATGTCACACCGAGGGGCTGGACCTCCCA 1756
DB 301 TCCAGCCCTGTCTAACCATCTCCAAATTAATGTCACACCGAGGGGCTGGACCTCCCA 360
QY 1757 CATCATTCATTTGCTTGTCTGCAAGTGGC 1785
DB 361 CATCATTCATTTGCTTGTCTGCAAGTGGC 389

RESULT 15
AAH12103/C
ID AAH12103 standard; cDNA; 578 BP.
AC AAH12103;
AC 26-JUN-2001 (first entry)
DT 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (3'-primer) SEQ ID NO:8938.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX
XX 27-AUG-1999; 99JP-00300253.
```

PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.

XX  
XX  
XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

XX  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.

XX  
XX  
XX Claim 3; SEQ ID NO 8938; 2537bp + Sequence Listing; English.

XX  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dr primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX

XX  
XX  
XX Sequence 578 BP; 134 A; 142 C; 168 G; 127 T; 0 U; 7 Other;

XX  
XX  
XX Query Match 18.9%; Score 341.8; DB 4; Length 578;  
XX Best Local Similarity 93.6%; Pred. No. 5.3e-82;

XX  
XX  
XX Matches 396; Conservative 0; Mismatches 23; Indels 4; Gaps 4;

QY 1392 GCTCAGTACGATTTGTCGTCATCTACCCACAGACGAGTGGGCCACCCAGAACCA 1451

DB 576 GCTCAGTACGATTTGTCGTCATCTACCCACAGACGAGTGGGCCACCCAGAACCA 517

QY 1452 AGCCCCCTTGAATGATATCTTTTCATCA-GGGTTGCTATGGGGCCAGGGCAGAGGT 1510

DB 516 AGCCCCCTTGAATGATATCTTTTCATCA-GGGTTGCTATGGGGCCAGGGCAGAGGT 457

QY 1511 ATGGCCCTTGGC-AGGGTAGAGA-GATTATCAACCAGGAAACCCAGTATTAAAG 1568

DB 456 TAGGCCCTTGGCAGAGGTAGAGACATTATCAACCCAGGAAACCCAGTATTAAAG 397

QY 1569 AAGCCCCCTTGGGGGAGAGACATATAGCAGGGGTGGGCTGCTTCTTATCTTGAC 1628

DB 396 AAGCCCCCTTGGGGGAGAGACATATAGCAGGGGTGGGCTGCTTCTTATCTTGAC 338

QY 1629 AATCTAGTCGATTTCTGCTTTTCTCCGATTTGGGGATTTGGGGGACCACTTAAGA 1688

DB 337 AATCTAGTCGATTTCTGCTTTTCTCCGATTTGGGGATTTGGGGGACCACTTAAGA 278

QY 1689 TGGCTCTTCCAGCCCTGTCTCAACCACTTCCAAATTAAGTGCACCCAGGGGCTTGGC 1748

DB 277 TGGCTCTTCCAGCCCTGTCTCAACCACTTCCAAATTAAGTGCACCCAGGGGCTTGGC 218

QY 1749 ACCTCCACATCATTCATTGTCTTGTGCGCAAGTGCATTAACGGCGTGTATTGCCAAC 1808

DB 217 ACCTCCACATCATTCATTGTCTTGTGCGCAAGTGCATTAACGGCGTGTATTGCCAAC 158

QY 1809 TGG 1811

DB 157 TGG 155

Search completed: March 28, 2005, 01:58:00  
Job time : 1020 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2005, 01:39:39 ; Search time 323 Seconds  
(without alignments)  
9174.295 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtgccttcgagccctcgcgcg.....cgcgctgattgcacacccg 1811

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405566

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTCUTS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.4	3.0	17930	4	US-09-949-016-11902
2	54.4	3.0	18351	4	US-09-949-016-16749
3	54.2	3.0	60489	4	US-09-949-016-16287
4	52.8	2.9	455726	4	US-09-949-016-14157
5	52.8	2.9	481115	4	US-09-949-016-11940
6	51.4	2.8	199471	4	US-09-949-016-14083
7	49.8	2.7	174493	4	US-09-804-471A-3
8	49.8	2.7	174493	4	US-10-238-709-3
9	49.6	2.7	133871	4	US-09-949-016-11863
10	48.6	2.7	161607	4	US-09-949-016-12210
11	48.4	2.7	601	4	US-09-949-016-50874
12	48.4	2.7	113701	4	US-09-949-016-13214
13	48.2	2.7	26845	4	US-09-949-016-11815
14	48.2	2.7	27132	4	US-09-949-016-15424
15	48.2	2.7	278866	4	US-09-949-016-13922
16	48.2	2.7	278866	4	US-09-949-016-13923
17	48.2	2.7	278866	4	US-09-949-016-13924
18	48.2	2.7	278866	4	US-09-949-016-13925
19	48.2	2.7	278866	4	US-09-949-016-13926
20	48.2	2.7	278866	4	US-09-949-016-14699
21	48.2	2.7	278866	4	US-09-949-016-14700
22	48.2	2.7	278866	4	US-09-949-016-14701
23	48.2	2.7	278866	4	US-09-949-016-14702
24	48.2	2.7	278866	4	US-09-949-016-14703
25	48	2.7	331814	4	US-09-949-016-12008
26	48	2.7	331814	4	US-09-949-016-17056
27	47.4	2.6	26760	4	US-09-949-016-15894

c	28	46.2	2.6	30001	1	US-08-125-468-1	Sequence 1, Appl1
c	29	46.2	2.6	30001	2	US-08-474-933-1	Sequence 1, Appl1
c	30	45.8	2.5	103750	4	US-09-949-016-113319	Sequence 13319, A
	31	45.6	2.5	601	4	US-09-949-016-134573	Sequence 134573, A
	32	45.6	2.5	601	4	US-09-949-016-134574	Sequence 134574, A
	33	45.6	2.5	601	4	US-09-949-016-134575	Sequence 134575, A
	34	45.6	2.5	43375	4	US-09-949-016-12688	Sequence 12688, A
	35	45.6	2.5	43376	4	US-09-949-016-15515	Sequence 15515, A
	36	45.6	2.5	72928	3	US-09-009-913-1	Sequence 1, Appl1
	37	45.4	2.5	601	4	US-09-949-016-166485	Sequence 166485, A
	38	45.4	2.5	30054	4	US-09-949-016-16429	Sequence 16429, A
	39	45.4	2.5	35122	4	US-09-949-016-11873	Sequence 11873, A
c	40	45.4	2.5	187136	4	US-09-949-016-17231	Sequence 17231, A
	41	45.2	2.5	247299	4	US-09-949-016-17590	Sequence 17590, A
	42	45	2.5	374159	4	US-09-949-016-15868	Sequence 15868, A
	43	44.4	2.5	262	4	US-09-573-080A-99	Sequence 99, Appl1
c	44	44.4	2.5	165651	4	US-09-949-016-13032	Sequence 13032, A
	45	44.2	2.4	13945	4	US-09-949-016-14231	Sequence 14231, A

## ALIGNMENTS

```
RESULT 1
US-09-949-016-11902
; Sequence 11902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C100107
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11902
; LENGTH: 17930
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11902

Query Match          3.0%; Score 54.4; DB 4; Length 17930;
Best Local Similarity 70.2%; Pred. No. 0.00029;
Matches 73; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 487 TGACCTTCACTTCAGTTTGTGATCCGTAATAATGACAATTCGAAGCTTACAGTGC 546
DB 13712 TCACGCGAGCTTCATTTCTCATTTGTTAATGAGCGATTAGACCTCAAGGAC 13771
QY 547 TGTGAGAGATTAATGAATCAATCTGTTGAAGCTTTGCA 590
DB 13772 TGTGTAAGATGAATGAATCAATCAATGTAAGTTCTTACA 13815

RESULT 2
US-09-949-016-16749
; Sequence 16749, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C100107
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 16749  
; LENGTH: 18351  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16749

Query Match 3.0%; Score 54.4; DB 4; Length 18351;  
Best Local Similarity 70.2%; Pred. No. 0.0003;  
Matches 73; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 487 TGACTTCACTGAGTTTGTGATCCGTAATAATGAGCAAAATTCGAACTTCACTGAGTGC 546  
DB 13712 TCACTGAGCTTCTTTCTCATTTGTTAATGAGGCTATTAGGACCTCACTCAAGAGC 13771  
QY 547 TGTGAGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGCA 590  
DB 13772 TGTGTAAGATGAATAATCAATACATGTAAGCTTTAACA 13815

RESULT 3  
US-09-949-016-16287/c  
; Sequence 16287, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTUR, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 16287  
; LENGTH: 60489  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16287

Query Match 3.0%; Score 54.2; DB 4; Length 60489;  
Best Local Similarity 69.2%; Pred. No. 0.0006;  
Matches 74; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 490 CTTCACCTCAGTTTGTGATCCGTAATAATGAGCAAAATTCGAACTTCACTGAGTGC 549  
DB 12018 CTGTACTCTGTTTCTCCTCATGTGTAACAAGGATTAAGAACTTCACTGAGTGC 11959  
QY 550 TGAAGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGAGAGAGG 596  
DB 11958 TGTGAGATGAATCTATTATACATGTAAGCACTTTGAACAATG 11912

RESULT 4  
US-09-949-016-14157/c  
; Sequence 14157, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTUR, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 14157  
; LENGTH: 455726  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(455726)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14157

Query Match 2.9%; Score 52.8; DB 4; Length 455726;  
Best Local Similarity 61.8%; Pred. No. 0.0039;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGTGACTTCACTGAGTTTGTGATCCGTAATAATGAGCAAAATTCGAAAGC 533  
DB 38822 TGCCCTGGGAGGATTAATCTTCACTGAGTGTTCATCTGAACCATGTCAGTGGCC 38763  
QY 534 TACTTCACTGAGTGTGTTGAGAGATTAAATGAACAATGCTTTGTAAGCTCTTTGAGAG 593  
DB 38762 TGATTCAACAAGGTGTTGTAAGATTAACTGACATGCTGTGAAGAAGCACTTACAGAGA 38703  
QY 594 GCGAGCTCGGAAGCA 609  
DB 38702 TGCTTAGTCCAAAGCA 38687

RESULT 5  
US-09-949-016-11940/c  
; Sequence 11940, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTUR, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 11940  
; LENGTH: 48115  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(48115)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-11940

Query Match 2.9%; Score 52.8; DB 4; Length 48115;  
Best Local Similarity 61.8%; Pred. No. 0.004;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGTGACTTCACTGAGTTTGTGATCCGTAATAATGAGCAAAATTCGAAAGC 533  
DB 16211 TGCCCTGGGAGGATTAATCTTCACTGAGTGTTCATCTGAACCATGTCAGTGGCC 16152











	Query Match	Similarity	88.9%	Score	1610.2	DB	17	Length	1826
	Best Local	Similarity	99.8%	Pred.	No. 0				
	Matches	1612	Conservative	0	Mismatches	3	Indels	0	Gaps
Qy	197	AGTGCATGGATTCTGGGTTTTCGATATTCGCGGCGCGGTTCAACGTATGCTGTGCGGCTCC							256
Db	10	ATTCTCGGGTTCTGGGTTTCGATATTCGCGGCGCGGTTCAACGTATGCTGTGCGGCTCC							69
Qy	257	TTCGGGTGATGTCGTCGCGCGCGGATGCCCGGGAACGAGCTGACCGGGGATCTCCGGG							316
Db	70	TTCGGGTGATGTCGTCGCGCGCGGATGCCCGGGAACGAGCTGACCGGGGATCTCCGGG							129

QY 317 CCCGAGCATTCGCGGCTGCAGATTGACGGGGATCCCGGATGACCCGCGCGCCCGCCG 376  
 Db 130 CCCGAGCATTCGCGGCTGCAGATTGACGGGGATCCCGGATGACCCGCGCGCCCGCCG 189  
 QY 377 CCTGACCGACGGGTTCAGACCTGGTGGGAAAGAGTGGGGAGCGGTCCTGAGATTC 436  
 Db 190 CCTGACCGACGGGTTCAGACCTGGTGGGAAAGAGTGGGGAGCGGTCCTGAGATTC 249  
 QY 437 CGATGCTTACGAGCGCAAGATGCTCACTTTATAGGTGACCTTACACATGTGACTTACC 496  
 Db 250 CGATGCTTACGAGCGCAAGATGCTCACTTTATAGGTGACCTTACACATGTGACTTACC 309  
 QY 497 TCAGTTTGTGATCCGTAAATGGAACAATCGAAGCTACTTCAAGGCTGTGAGAGG 556  
 Db 310 TCAGTTTGTGATCCGTAAATGGAACAATCGAAGCTACTTCAAGGCTGTGAGAGG 369  
 QY 557 ATTAATGAAAACAATGCTTTGTAAGCTCTTTCAGAGAGGAGCTCGAAAGCAGGGCTG 616  
 Db 370 ATTAATGAAAACAATGCTTTGTAAGCTCTTTCAGAGAGGAGCTCGAAAGCAGGGCTG 429  
 QY 617 GCCCGCAGAGACACCTGCTGTCAACAGGACCAAGGAGCATGAAAGATCCCGTGGAG 676  
 Db 430 GCCCGCAGAGACACCTGCTGTCAACAGGACCAAGGAGCATGAAAGATCCCGTGGAG 489  
 QY 677 CTGCGCCGTCAAGTGGAGTGCAGACCTCGGCTTCAGACAGCGCTGCGAGGTGCTTACCG 736  
 Db 490 CTGCGCCGTCAAGTGGAGTGCAGACCTCGGCTTCAGACAGCGCTGCGAGGTGCTTACCG 549  
 QY 737 GTCAAGGCCAGAGCATGATATGTGATGAGACTCTGTTTGGCAGCCAGCAGGACCCGG 796  
 Db 550 GTCAAGGCCAGAGCATGATATGTGATGAGACTCTGTTTGGCAGCCAGCAGGACCCGG 609  
 QY 797 CTTATCCCAACCGGACTTGCATCCGCTTGGGTGAGAGAGCTTAAACAACAAGAGGCTG 856  
 Db 610 CTTATCCCAACCGGACTTGCATCCGCTTGGGTGAGAGAGCTTAAACAACAAGAGGCTG 669  
 QY 857 GGCAGAGAGCATGAGAGGCTTGGGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAG 916  
 Db 670 GGCAGAGAGCATGAGAGGCTTGGGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAG 729  
 QY 917 GGCAGACACCCCACTTCAACAAGAGAGAAACAATACAGACCCATCAGCCACACC 976  
 Db 730 GGCAGACACCCCACTTCAACAAGAGAGAAACAATACAGACCCATCAGCCACACC 789  
 QY 977 CCGCTTACTGTATGATGCTGCTGTTGGCTCCGATCTGAAGCGCCACCTTCGGGGCC 1036  
 Db 790 CCGCTTACTGTATGATGCTGCTGTTGGCTCCGATCTGAAGCGCCACCTTCGGGGCC 849  
 QY 1037 CCGCGATGCGAAGGGGGATGCGCAAGACTCCGCTCTCTTGTGAGACGCAACCACT 1096  
 Db 850 CCGCGATGCGAAGGGGGATGCGCAAGACTCCGCTCTCTTGTGAGACGCAACCACT 909  
 QY 1097 ACCCCAGGGGTAGCCACTGCGCCCGCCCAAGGAGCACTGCGAGCCATTCACCCA 1156  
 Db 910 ACCCCAGGGGTAGCCACTGCGCCCGCCCAAGGAGCACTGCGAGCCATTCACCCA 969  
 QY 1157 GCTGATCCCTCAAGAGAGCGGGGCAAGCGCAGACTCCAGAGTATCTAATGGGT 1216  
 Db 970 GCTGATCCCTCAAGAGAGCGGGGCAAGCGCAGACTCCAGAGTATCTAATGGGT 1029  
 QY 1217 GGGTTACACTTTCACGCCCCCTGAAAGCGGGGACTTTCCTCATTCCTCACCCACTGAAT 1276  
 Db 1030 GGGTTACACTTTCACGCCCCCTGAAAGCGGGGACTTTCCTCATTCCTCACCCACTGAAT 1089  
 QY 1277 GTCCCCAGACCTGCTATTCAGCCACAGTGCCTCCCAACAATGGGGCTTCAGATTC 1336  
 Db 1090 GTCCCCAGACCTGCTATTCAGCCACAGTGCCTCCCAACAATGGGGCTTCAGATTC 1149  
 QY 1337 AGGCTTTCAGAGTCAAGGGGTGACCTTCGAGAGCCCTGTGATCTTCAGAGGGCTGCTCA 1396  
 Db 1150 AGGCTTTCAGAGTCAAGGGGTGACCTTCGAGAGCCCTGTGTGATCTTCAGAGGGCTGCTCA 1209

QY 1397 GTTAGCATTTCAAGCATCTACCCACGACGAGTGGGGCCCAACCAAGAAACCAAGCCC 1456  
 Db 1210 GTTAGCATTTCAAGCATCTACCCACGACGAGTGGGGCCCAACCAAGAAACCAAGCCC 1269  
 QY 1457 CTTGGAATATGATATCTTTTCATCAGGGTTGCTTAATGGGGCCACGCGCAGAGTATGGCC 1516  
 Db 1270 CTTGGAATATGATATCTTTTCATCAGGGTTGCTTAATGGGGCCACGCGCAGAGTATGGCC 1329  
 QY 1517 CTTTCCAGGGGTGAGAGACATTCATCACCAGGAAACCCAGATATTAAGAAAGCCCT 1576  
 Db 1330 CTTTCCAGGGGTGAGAGACATTCATCACCAGGAAACCCAGATATTAAGAAAGCCCT 1389  
 QY 1577 GTGGGGGACAGACAGATAGCAGGGGTGGGCAAGTGCCTCCCTTATCTGACATCTTA 1636  
 Db 1390 GTGGGGGACAGACAGATAGCAGGGGTGGGCAAGTGCCTCCCTTATCTGACATCTTA 1449  
 QY 1637 GTGATTTCTTCCCTTTTCTCCGATTCGGAATTTGGGGGCCACTTAAGATGCTCTC 1696  
 Db 1450 GTGATTTCTTCCCTTTTCTCCGATTCGGAATTTGGGGGCCACTTAAGATGCTCTC 1509  
 QY 1697 TCCAGCCCTGTTCACACATCTCCAAATTAATGTCACACCCAGGGGCTGGCACTTCCA 1756  
 Db 1510 TCCAGCCCTGTTCACACATCTCCAAATTAATGTCACACCCAGGGGCTGGCACTTCCA 1569  
 QY 1757 CATCATTCATTTCTTGTGTCGCAAGTGCAGATTAAGGGCTGATTCGCAACTGG 1811  
 Db 1570 CATCATTCATTTCTTGTGTCGCAAGTGCAGATTAAGGGCTGATTCGCAACTGG 1624

RESULT 2  
 US-10-452-858C-78  
 ? Sequence 78; Application US/10452858C  
 ? Publication No. US20040086945A1  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Streekristhna, Kocikanyadanam  
 ? APPLICANT: Gerwe, Gina S.  
 ? APPLICANT: Toerner, Daniel R.  
 ? TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THEREFOR  
 ? FILE REFERENCE: 8956P  
 ? CURRENT APPLICATION NUMBER: US/10/452, 858C  
 ? CURRENT FILING DATE: 2003-06-02  
 ? NUMBER OF SEQ ID NOS: 94  
 ? SOFTWARE: PatentIn version 3.1  
 ? SEQ ID NO 78  
 ? LENGTH: 709  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 ? FEATURE:  
 ? NAME/KEY: CDS  
 ? LOCATION: (21)..(707)  
 ? OTHER INFORMATION:  
 ? FEATURE:  
 ? NAME/KEY: misc.feature  
 ? LOCATION: (670)..(695)  
 ? OTHER INFORMATION: n represents a, c t or g  
 US-10-452-858C-78

Query Match 38.0%; Score 688.2; DB 17; Length 709;  
 Best Local Similarity 99.1%; Pred. No. 5.8e-200;  
 Matches 690; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 682 CGTCACTGGAGATGACAGACCTCGGCTTTCAGACCGGCTGCGAGGTGCTACCGGGTCAA 741  
 Db 14 CGTCAAGGAGATGACAGACCTCGGCTTTCAGACCGGCTGCGAGGTGCTACCGGGTCAA 73  
 QY 742 GGCAGAGAGTATATGATGATGAGACTCTGTTTGGCAGCCGACGAGGACCCCGGCTAC 801  
 Db 74 GGCAGAGAGTATATGATGATGAGACTCTGTTTGGCAGCCGACGAGGACCCCGGCTAC 133  
 QY 802 CCCACCGGACTTTCGATCCGCTTGGGTGAGAGAGGTAAACAGAACCAAGAGGGGTGGGCAA 861  
 Db 134 CCCACCGGACTTTCGATCCGCTTGGGTGAGAGAGGTAAACAGAACCAAGAGGGGTGGGCAA 193



```

Qy 862 GAGAGCATGAAAGCCTTGGGGGCAAGGGGAGCTGTAGACACCCCTCAAGGGGAG 921
Db 194 GGAGGCACTGAAAGCCTTGGGGGCAAGGGGAGCTGTAGACACCCCTCAAGGGGAG 253
Qy 922 CACCCCTCACTCAACCAAGGAGAAACAATACAGACCCATCAGCCACACCCCGTC 981
Db 254 CACCCCTCACTCAACCAAGGAGAAACAATACAGACCCATCAGCCACACCCCGTC 313
Qy 982 TTACTGTAGTAGTGGCTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCCCGG 1041
Db 314 TTACTGTAGTAGTGGCTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCCCGG 373
Qy 1042 GATGCGAAGGGGAGTGCAGCAAGCTCCGCTCTCTTGTGAGCGCACTACCC 1101
Db 374 GATGCGAAGGGGAGTGCAGCAAGCTCCGCTCTCTTGTGAGCGCACTACCC 433
Qy 1102 CAGGGGTAGCCACTGCGCCCGCCCAAGGAGGACACCTGCGAGCCATTACCCAGCTGG 1161
Db 434 CAGGGGTAGCCACTGCGCCCGCCCAAGGAGGACACCTGCGAGCCATTACCCAGCTGG 493
Qy 1162 TCCCTCAAGAGAGAGCGGGGCGAGCGGAGACCTCCAGAGTTATCTAGGGTGGCTT 1221
Db 494 TCCCTCAAGAGAGAGCGGGGCGAGCGGAGACCTCCAGAGTTATCTAGGGTGGCTT 553
Qy 1222 ACACCTCTTACGCGCCCTGAGAGCGGGGAGCTTCCCATTCCTCACCACCTGAATGTCCC 1281
Db 554 ACACCTCTTACGCGCCCTGAGAGCGGGGAGCTTCCCATTCCTCACCACCTGAATGTCCC 613
Qy 1282 CAGCACTGGTCACTCAAGCCACCAAGTGCAGGCGCCCAACAATGGGCTCAGAGATCTCAGGCC 1341
Db 614 CAGCACTGGTCACTCAAGCCACCAAGTGCAGGCGCCCAACAATGGGCTCAGAGATCTCAGGCC 673
Qy 1342 TTCGAGCTCAGGGGTGAGCTTCCGAGAGCGCCCTGTGT 1377
Db 674 TTCGAGCTCAGGGGTGAGCTTCCGAGAGCGCCCTGTGT 709

```

## RESULT 3

```

US-09-918-995-1667
; Sequence 1667, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1667
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(477)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-1667

```

## Query Match

```

24.1%; Score 436.6; DB 10; Length 477;
Best Local Similarity: 99.1%; Pred. No. 5.9e-123;
Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 986 TGTGATGATGAGCTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCCGGGAGT 1045
Db 35 TGTGATGATGAGCTTTGGCTCCGATCTGAAGGCGCAGCTTGGGGGCGCCGGGAGT 94
Qy 1046 GCGAAGGGGAGTGCAGCAAGCTCCGTCTCTTGTGAGCGCCACCACTTACCCCAAG 1105
Db 95 GCGAAGGGGAGTGCAGCAAGCTCCGTCTCTTGTGAGCGCCACCACTTACCCCAAG 154

```

```

Qy 1106 GGTAGCACTGCGCCCGCCGAGGGGACCACTTGCAGACCTTCACCAGCTGTCCC 1165
Db 155 GGTAGCACTGCGCCCGCCGAGGGGACCACTTGCAGACCTTCACCAGCTGTCCC 214
Qy 1166 TCCAAAGAGAGCCGGGCGCAGCGGAGACTTCCAGAAATTATCTATGGGTGGTTACAC 1225
Db 215 TCCAAAGAGAGCCGGGCGCAGCGGAGACTTCCAGAAATTATCTATGGGTGGTTACAC 274
Qy 1226 TTTTCAAGCCCCCTGAAGCGGGGACTTTCCATTCTCTCAGCCAGCTGAATGTCCCAGC 1285
Db 275 TTTTCAAGCCCCCTGAAGCGGGGACTTTCCATTCTCTCAGCCAGCTGAATGTCCCAGC 334
Qy 1286 ACTGTCATTCAGGACGACAGTGCAGGAGCCCGCAACAATGGGCTCAGAGATCTCAGGCTTCC 1345
Db 335 ACTGTCATTCAGGACGACGAGTGCAGGAGCCCGCAACAATGGGCTCAGAGATCTCAGGCTTCC 394
Qy 1346 AGTCAGGGGTGAGCTTCCGAGGCGCCCTGTGATCTTCCAGGGCTCGCTCAGTTAGCAAT 1405
Db 395 AGTCAGGGGTGAGCTTCCGAGGCGCCCTGTGATCTTCCAGGGCTCGCTCAGTTAGCAAT 454
Qy 1406 TCAGTCCATCTTACCCCAAGAG 1428
Db 455 TCAGTCCATCTTACCCCAAGAG 477

```

## RESULT 4

```

US-09-918-995-31166
; Sequence 31166, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31166
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(484)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31166

```

## Query Match

```

22.7%; Score 410.6; DB 10; Length 484;
Best Local Similarity: 93.9%; Pred. No. 5.4e-115;
Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

```

Qy 426 CTTGAGGATCCCGATGCTTAGAGCCAAAGATGCTCAGCTTTATAGTGTGACTTACAT 485
Db 39 CTTGAGGATCCCGATGCTTAGAGCCAAAGATGCTCAGCTTTATAGTGTGACTTACAT 98
Qy 486 GTGACTTCACTCAGCTTTGTGATCCGTAATGAGCAAAATGCAATTCGAAGCTTACAGAGT 545
Db 99 GTGACTTCACTCAGCTTTGTGATCCGTAATGAGCAAAATGCAAAAGCTTACAGAGT 158
Qy 546 CTGTTGAGAGGATTAATGAACAATGCTTTGAAAGCTTTTGCAGAGGAGAGCTTCGGA 605
Db 159 CTGTTGAGAGGATTAATGAACAATGCTTTGAAAGCTTTTGCAGAGGAGAGCTTCGGA 218
Qy 606 AGCAGGGCTTGGCGGCGAGACACACTGTGTGACCAAGGAGACCAAGGAGCATGAAGA 665
Db 219 AGCAGGGCTTGGCGGCGAGACACACTGTGTGACCAAGGAGACCAAGGAGCATGAAGA 278
Qy 666 CCCCCTGAGACTGAGCGGCTCAGTGGAGTGCAGACCTTCGAGCTTCAAGCAGCGCTGCGAG 725

```

Db 279 CCCCCGTGAGCTGGCCCTCATGAGGATGACAGCCCTTGCCCTTCAACACGCTGCCGAG 338  
Qy 726 GTGGCTACCGGGTAAAGCCGACGTCATATGTGATGATGATCTCTGTTGGCAGCCAG 785  
Db 339 GTGGCTACCGGGTAAAGCCGACGTCATATGTGATGATGATCTCTGTTGGCAGCCAG 398  
Qy 786 CAGGACACCGGCTACCCCAACCGGACTTCGATCCGCCCTGGGTGAGAGAGCTTAACAGAA 845  
Db 399 CAGGACACCGGCTACCCCAACCGGACTTCGATCCGCCCTGGGTGAGAGAGCTTAACAGAA 458  
Qy 846 CCAAGAGCGGTGGGCAAGAGGAGCATCG 871  
Db 459 CCAGAGGCGGTGGGCAAGAGGAGCATCG 484

## RESULT 5

US-09-864-408A-8425  
; Sequence 8425, Application US/09864408A  
; Publication No. US20040009474A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard A.  
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encc  
; FILE REFERENCE: 21402-012  
; CURRENT APPLICATION NUMBER: US/09/864,408A  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/206,690  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 9068  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 8425  
; LENGTH: 292  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-864-408A-8425

Query Match 7.2%; Score 130.8; DB 11; Length 292;  
Best Local Similarity 72.8%; Pred. No. 2.3e-29;  
Matches 182; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

Qy 639 CACGAGGAGCACGACGATGAAAGACCCCTGAG-AGCTGGCCGTCATGGGATGACAG 697  
Db 43 CACGACACCCACTGGGACGATGAAGACCCCGTGGGACCGGCCACGACGAGATATAA 102  
Qy 698 ACCCTGGGCTTACGACACCGCTGCGGAGTGGCTACACGGGTCAAGGCGAGAGCTCAT 757  
Db 103 GCATTCACCTTACGACACCGGATCCCAACGATCAAGGCTCAAGGCGAGATCTAT 162  
Qy 758 GTGATGAGACTGTTTGGCAGCCGACGAGGACCCGCTACCCCAACGAGCTTCGAT 817  
Db 163 GTGATGAGACTGTTTGGCAGCCGACGAGGACCCGCTACCCCAACGAGCTTCGAT 222  
Qy 818 CCGGCTGGGTGGGAGAGGCTTAACAGAACGAGGCGCTGGGCAAGAGGATCGAAGGCC 877  
Db 223 CCACCTGGGTGGGAGAGGATTAACGATCCAGAGAGAGTGGGCGCCACGAGACCGAAGGCG 282  
Qy 878 TTGGGGGCA 887  
Db 283 TCTCTGGCA 292

## RESULT 6

US-09-908-975-4892  
; Sequence 4892, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, Avi  
; APPLICANT: WASSERMAN, Alon  
; APPLICANT: MINTZ, Eli  
; APPLICANT: MINTZ, Ilat  
; APPLICANT: PALGLER, Simchon  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC  
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME

FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/287,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 4892  
; LENGTH: 60  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-975-4892

Query Match 3.3%; Score 60; DB 10; Length 60;  
Best Local Similarity 100.0%; Pred. No. 7e-08;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 164 TGCTGCTGCTGCTCCCTGGTTCGAGTGCAGAAAGTGTGGGTTCTGGGTTCTGGATTC 223  
Db 1 TGCTGCTGCTGCTCCCTGGTTCGAGTGCAGAAAGTGTGGGTTCTGGGTTCTGGATTC 60

## RESULT 7

US-10-027-632-288030  
; Sequence 288030, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108427.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 288030  
; LENGTH: 479  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-288030

Query Match 2.9%; Score 52.8; DB 13; Length 479;  
Best Local Similarity 61.8%; Pred. No. 2e-05;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 474 TGACCTACACATGTGACTTACCTGATTTTGTGATCCGTAAATGACAAATTCGAGC 533  
Db 89 TGGCCGTGGGAGGTAATCTACCTGATTTTGTGATCCGTAAATGACAAATTCGAGC 148  
Qy 534 TACTTACAGTGTGTTGAGAGATTAATGAACAGTCTGTAAGCTCTTTGCAGGA 593  
Db 149 TGAATCAAGGTGTGTTGAAGATTAATGACATGACGTGTGAAGAGCACTTAAGGAA 208  
Qy 594 GGAAGCTCGGAAGCA 609  
Db 209 TGCTTAGTCCAAAGCA 224

```
RESULT 8
US-10-027-632-288030
; Sequence 288030, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 288030
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288030

Query Match      2.9%; Score 52.8; DB 17; Length 479;
Best Local Similarity 61.8%; Pred. No. 2e-05;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGTGACTTCACTTCACTTTTGTGATCCGTAATAATGACAAATTGCAAGC 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 89 TGGCTGGGGAGGTAACTTCACTTCACTTTTGTGATCCGTAATAATGACAAATTGCAAGC 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 534 TACTTCACAGTCTGTGAGAGATTAAATGAACAATGCTTTGCGAGA 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 149 TGATTCAACAGGTGTTGTAAGATTAACTGACATGACGTGTAAGAGACATTAGCAGAA 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 594 GGGAGCCTCGAAGCA 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 TGCTTAGTCAAAAGCA 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-10-719-993-6787/c
; Sequence 6787, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6787
; LENGTH: 493999
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(493999)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6787
```

```
Query Match      2.9%; Score 52.8; DB 18; Length 493999;
Best Local Similarity 61.8%; Pred. No. 0.00016;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGTGACTTCACTTCACTTTTGTGATCCGTAATAATGACAAATTGCAAGC 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71520 TGGCTGGGGAGGTAACTTCACTTCACTTTTGTGATCCGTAATAATGACAAATTGCAAGC 71461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 534 TACTTCACAGTCTGTGAGAGATTAAATGAACAATGCTTTGCGAGA 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71460 TGATTCAACAGGTGTTGTAAGATTAACTGACATGACGTGTAAGAGACATTAGCAGAA 71401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 594 GGGAGCCTCGAAGCA 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71400 TGCTTAGTCAAAAGCA 71385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-10-242-355-1101
; Sequence 1101, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1101
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1101

Query Match      2.8%; Score 51.6; DB 17; Length 680;
Best Local Similarity 67.9%; Pred. No. 5.3e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTGACCTCAGTTTGTGATCCGTAATAATGACAAATTGCAAGCTTCACTTCAAGTCTGT 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 35 CTTGACCTCAGTTTGTGATCCGTAATAATGACAAATTGCAAGCTTCACTTCAAGTCTGT 94
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 550 TGAGAGATTAAATGAACAATGCTTTGTAAGCTTTGCGAGAG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 TAAAGATTAAATGAATTAATCCATGAAAGCTTGTAGCAGATG 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-10-242-355-1102
; Sequence 1102, Application US/10242355
```

```
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1102
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1102

Query Match      2.8%; Score 51.6; DB 17; Length 680;
Best Local Similarity 67.9%; Pred. No. 5.3e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY      490 CTTCACTCAGTTTGGATCCGTAATAATGCAAAATTCGAAGCTTCTACAGTGTGT 549
Db      35 CCTTACCTCAATCTCCTTATCTGTGAATGCGAATATAAATCTACCTACAGAAATTAT 94

QY      550 TGAGAGATTAAATGAACAATGCTTTAAAGCTCTTGAGAGAGG 595
Db      95 TAACAGATTAAATGAGATTATTCATGGAAGCTAGTAGAGCATG 140

RESULT 12
US-10-242-355-240
; Sequence 240, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
```

```
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 240
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-240

Query Match      2.8%; Score 50.8; DB 17; Length 343;
Best Local Similarity 66.0%; Pred. No. 7.6e-05;
Matches 70; Conservative 2; Mismatches 34; Indels 0; Gaps 0;

QY      490 CTTCACTCAGTTTGGATCCGTAATAATGCAAAATTCGAAGCTTCTACAGTGTGT 549
Db      35 CCTTACCTCAATCTCCTTATCTGTGAATGCGAATATAAATCTACCTACAGAAATTAT 94

QY      550 TGAGAGATTAAATGAACAATGCTTTAAAGCTCTTGAGAGAGG 595
Db      95 TAACAGATTAAATGAGATTATTCATGGAAGCTAGTAGAGCATG 140

RESULT 13
US-10-027-632-102308/c
; Sequence 102308, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102308
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102308

Query Match      2.8%; Score 50.4; DB 13; Length 741;
Best Local Similarity 60.0%; Pred. No. 0.00013;
Matches 84; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY      487 TGACTTCACTCAGTTTGGATCCGTAATAATGCAAAATTCGAAGCTTCTACAGTGT 546
Db      444 TACTTAAGACACAGTTTCTTTATCTGTAAATAGTACATGAATGACATCTTGAAAGGCG 385

QY      547 TGTGAGAGATTAAATGAACAATGCTTTAAAGCTCTTTGAGAGAGGAGCTTGGA 606
Db      384 TACTGTGAGAGATTATGTGAATAATGAACGCAATGTATACAGAGTTCTGAGCTCAAAA 325
```

Qy	607	GCAGGGCCTGGCCGGCAGAG	626
Db	324	AGGCGAACTGTTATGATGAG	305

RESULT 14

```

US-10-027-632-102308/c
; Sequence 102308, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827, 129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102308
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-102308

```

Query Match	2.8%	Score 50.4;	DB 17;	Length 741;
Best Local Similarity	60.0%;	Pred. No. 0.00013;		
Matches	84;	Conservative	0;	Mismatches 56;
			Indels	0;
			Gaps	0

QY	487	TGACTTCACCTCAGTTTGTGATCCGTAAATGCAAAATTGCAGCTACTTACAGTGC	546
Db	444	TCACTAAGCACAAGTTCTTTATCTGTAAATAATGTACATGAAATACCATTTTGAAGGCC	385
QY	547	TGTTGAGAGGATTAATGAAACAATGCTTGAAGCTTTTGAGAGAGGAGCCTTGGAA	606
Db	384	TACTGTGAGATTATATGAAAATTAATTAACGCAATGTACTTACGAGGTTCTGCACTAAAA	325
QY	607	GCAGGAGCTTGCGCGGCAAGG	626
Db	324	AGGGCAACTGTTATGATGAG	305

## RESULT 15

US-10-292-798-1243  
Sequence 1243, Application US/10292798  
Publication No. US20030235833A1  
GENERAL INFORMATION:  
APPLICANT: SUMA, MAKIRO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTSUKA  
APPLICANT: AUBURATANI, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2070

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1243
; LENGTH: 43981
; MW: 2702
```

```

1 ORGANISM: Homo sapiens
2 FEATURE:
3 LOCATION: source
4 FEATURE:
5 LOCATION: (1)..(43981)
6 FEATURE:
7 NAME/KEY: CDS
8 LOCATION: (201)..(470)
9 FEATURE:
10 NAME/KEY: CDS
11 LOCATION: (12636)..(12748)
12 FEATURE:
13 NAME/KEY: CDS
14 LOCATION: (24946)..(25088)
15 FEATURE:
16 NAME/KEY: CDS
17 LOCATION: (32801)..(32997)
18 FEATURE:
19 NAME/KEY: CDS
20 LOCATION: (41097)..(41379)
21 FEATURE:
22 NAME/KEY: CDS
23 LOCATION: (43585)..(43781)
24 FEATURE:
25 NAME/KEY: modified_base
26 LOCATION: (2811)..(2910)
27 OTHER INFORMATION: a, t, c, g, unknown or other
28 FEATURE:
29 NAME/KEY: modified_base
30 LOCATION: (6914)..(6914)
31 OTHER INFORMATION: a, t, c, g, unknown or other
32 FEATURE:
33 NAME/KEY: modified_base
34 LOCATION: (6923)..(6923)
35 OTHER INFORMATION: a, t, c, g, unknown or other
36 FEATURE:
37 NAME/KEY: modified_base
38 LOCATION: (6932)..(6932)
39 OTHER INFORMATION: a, t, c, g, unknown or other
40 FEATURE:
41 NAME/KEY: modified_base
42 LOCATION: (6956)..(6956)
43 OTHER INFORMATION: a, t, c, g, unknown or other
44 FEATURE:
45 NAME/KEY: modified_base
46 LOCATION: (6964)..(6964)
47 OTHER INFORMATION: a, t, c, g, unknown or other
48 FEATURE:
49 NAME/KEY: modified_base
50 LOCATION: (7905)..(7905)
51 OTHER INFORMATION: a, t, c, g, unknown or other
52 JS-10-292-798-1243

```

Query Match	2.8%	Score 50.4;	DB 17;	Length 43981;
Best Local Similarity	71.7%	Pred. No. 0.00042;		
Matches 66;	Conservative 0;	Mismatches 26;	Indels 0;	Gaps 0;

QY	Db	QY	Db
495	CCGCAATTGTCATCCGTAATAATGCAAAATCGAAGTACTTCACATGCGTTGGAA	555	GGATTTAATGAAACAATGCTGTTAAAGCTTT
18915	CCGCAATTCCTCTATCTGTTTACTGTAAGAGATTAATCAACTTCTCAAAGGCTTTTGA	18975	GGATTTAATGAAATTAATTCGTAAAGCACTT

Mon Mar 28 09:43:46 2005

us-10-031-589-3.rmpb

Page 8

Search completed: March 28, 2005, 08:50:19  
Job time : 2119 secs

---



Result No.	Score	Query Match	Length	DB	ID	Description
1	1764.4	97.4	1766	3	CR595506	CR595506 full-1-eng
2	1764.4	97.4	1936	3	CR608011	CR608011 full-1-eng
3	1659	91.6	1839	3	CR623011	CR623011 full-1-eng
4	1631	90.1	1632	3	CR621376	CR621376 full-1-eng
5	1614	89.1	1614	3	CR621332	CR621332 full-1-eng
6	1600	88.1	1780	3	CR600183	CR600183 full-1-eng
7	1596	88.1	1596	3	CR597629	CR597629 full-1-eng
8	1557	85.1	1575	3	CR606890	CR606890 full-1-eng
9	1541	85.1	1560	3	CR628812	CR628812 full-1-eng
10	1346	74.3	1488	3	CR613777	CR613777 full-1-eng
11	1325	73.2	1467	3	CR622791	CR622791 full-1-eng
12	1323	73.1	1465	3	CR605763	CR605763 full-1-eng
13	1045	57.7	1045	3	CR598289	CR598289 full-1-eng
14	1012.2	55.9	1098	5	BX358473	BX358473 full-1-eng
15	976.8	53.9	1038	5	BX339727	BX339727 full-1-eng
16	971.2	53.6	1008	5	BX339726	BX339726 full-1-eng
17	932.4	51.5	1101	1	AL523765	AL523765 full-1-eng
18	930.4	51.4	1113	5	BX403928	BX403928 full-1-eng
19	924.4	51.0	1063	1	AL560267	AL560267 full-1-eng
20	909.2	50.2	951	5	BX354462	BX354462 full-1-eng
21	898.6	49.6	991	5	BX345478	BX345478 full-1-eng
22	892	49.3	892	3	CR596541	CR596541 full-1-eng
23	889.2	49.1	1006	1	AL582549	AL582549 full-1-eng
24	884.2	48.8	1047	5	BM903591	BM903591 full-1-eng

C	25	880.6	48.6	981	1	AL582286	AL582286
	26	880.6	48.6	1009	1	AL517881	AL517881
	27	876.8	48.4	1046	5	BX384471	BX384471
	28	876.6	48.4	1029	5	BX376800	BX376800
	29	862.2	47.6	1032	5	BX372860	BX372860
	30	860.2	47.5	1017	5	BM921038	BM921038
	31	857.6	47.4	972	1	AL561282	AL561282
	32	852	47.0	1063	1	AL560832	AL560832
	33	849	46.9	983	5	BX353787	BX353787
	34	845.8	46.7	926	5	BX3494189	BX3494189
C	35	835.4	45.1	872	5	BX346001	BX346001
	36	830.2	45.8	866	5	BX352684	BX352684
	37	826.8	45.7	950	5	BQ933060	BQ943060
	38	826.6	45.6	1047	1	AL581746	AL581746
C	39	823.4	45.5	855	5	BX366971	BX366971
	40	815	45.0	892	1	AL516514	AL516514
	41	809.6	44.7	879	5	BU526901	AGENCOCURT
	42	806	44.5	903	1	AL555180	AL555180
	43	803.4	44.4	907	5	BQ962573	AGENCOCURT
	44	791.6	43.7	1011	5	BX366455	BX366455
	45	785.8	43.4	1038	5	BX356915	BX356915

## ALIGNMENTS

## RESULT 1

CR595506

**DEFINITION** full-length cDNA clone CS0DG007YI18 of B cells (Ramos cell line) of *Homo sapiens* (human).

ACCESSION

## VERSLUITING

**SOURCE**

ORGANIT

.....

## REFERENCES

AULT  
TITTE

JOURNAL

REMARK

## REFERENCE

AUTHOR TITLE

ЛЛЕ  
ТОТЕНА

[illegible]

**COMMENT**

## FEATURES

sou

ORIGIN

?

## Query Part 1

Best  
Match

.....

Qy

2

DB

QY 75 GGAAGCTCTGAGGCTCAAGAGCGTGGGCGCCGCCCTGGGCGGAGCTCCCCCAATCCGCG 134  
 Db 61 GGAAGCTCTGAGGCTCAAGAGCGTGGGCGCCGCCCTGGGCGGAGCTCCCCCAATCCGCG 120  
 QY 135 GAGCGAATGGTCCGGGTCCGGTCCGAGTGTCTGTGGTCTCTCCGTGGTCTGGGTGC 194  
 Db 121 GAGCGAATGGTCCGGGTCCGGTCCGAGTGTCTGTGGTCTCTCCGTGGTCTGGGTGC 180  
 QY 195 AAGATGCTGGGTCTGGGTCTGTGGATTCGGGGCCGTTCAAGATGAGCTGTGCGGCT 254  
 Db 181 AAGATGCTGGGTCTGGGTCTGTGGATTCGGGGCCGTTCAAGATGAGCTGTGCGGCT 240  
 QY 255 CTTGGGAGTCCGTCGCGCGCGGTGCGCCGCGGAGCGGCTAGGCTGCGGGGGTCCG 314  
 Db 241 CTTGGGAGTCCGTCGCGCGCGGTGCGCCGCGGAGCGGCTAGGCTGCGGGGGTCCG 300  
 QY 315 GAGCCCAAGGATTCGGGCTGCAAGTTGACGGGATCCCGATGCAACCGCGCCCGC 374  
 Db 301 GAGCCCAAGGATTCGGGCTGCAAGTTGACGGGATCCCGATGCAACCGCGCCCGC 360  
 QY 375 GCGCTCACGAGCGGTCCAGACTGTGGGAAAGGTGCGGGGACGGGTCCCTGAGAT 434  
 Db 361 GCGCTCACGAGCGGTCCAGACTGTGGGAAAGGTGCGGGGACGGGTCCCTGAGAT 420  
 QY 435 CCGCATGCTACGAGCCAAAGTCTGAGTTATAGGTGACCTTACATGATGATCTCA 494  
 Db 421 CCGCATGCTACGAGCCAAAGTCTGAGTTATAGGTGACCTTACATGATGATCTCA 480  
 QY 495 CTTGATGTTTGTGATCCGTAAATGACAATTTGAACTTACATGATGATGATGAGA 554  
 Db 481 CTTGATGTTTGTGATCCGTAAATGACAATTTGAACTTACATGATGATGATGAGA 540  
 QY 555 GGAATTAATGAACAAATGCTTTTAAAGCTCTTTGCAAGAGGAGGCTCGGAAAGAGGCC 614  
 Db 541 GGAATTAATGAACAAATGCTTTTAAAGCTCTTTGCAAGAGGAGGCTCGGAAAGAGGCC 600  
 QY 615 TGGCGGCGAAGCACTGCTGTCAACAGGAGCAACAGGAGCATGAAAGAGGAGGCC 674  
 Db 601 TGGCGGCGAAGCACTGCTGTCAACAGGAGCAACAGGAGCATGAAAGAGGAGGCC 660  
 QY 675 AGCTGCGGCTGAGTGGGATGACAGCTCGGCTTTCAGCAACCGGTGCGAGGTGCTACC 734  
 Db 661 AGCTGCGGCTGAGTGGGATGACAGCTCGGCTTTCAGCAACCGGTGCGAGGTGCTACC 720  
 QY 735 GGGTCAAGGCGCAAGAGCTCATATGTGATGAGACTCTTTTGGAGGCCAGAGGACCC 794  
 Db 721 GGGTCAAGGCGCAAGAGCTCATATGTGATGAGACTCTTTTGGAGGCCAGAGGACCC 780  
 QY 795 GAGCTACCCCAAGGAGCTTGTGATCCGCTGGGTGAGAGGAGGCTTAAAGAGAGGAG 854  
 Db 781 GAGCTACCCCAAGGAGCTTGTGATCCGCTGGGTGAGAGGAGGCTTAAAGAGAGGAG 840  
 QY 855 TGGGCAAGAGGAGCTTGGAGGCTTGGGAGCAAGGAGAGCTGTGAGACCAACCCCTCAA 914  
 Db 841 TGGGCAAGAGGAGCTTGGAGGCTTGGGAGCAAGGAGAGCTGTGAGACCAACCCCTCAA 900  
 QY 915 GGGGCAAGAGGAGCTTGTGATCCGCTGGGTGAGAGGAGGCTTAAAGAGAGGAGG 974  
 Db 901 GGGGCAAGAGGAGCTTGTGATCCGCTGGGTGAGAGGAGGCTTAAAGAGAGGAGG 960  
 QY 975 CCGGCTTGTGATGAGTGGGTGTTGGCTCCGATCTGAAAGGAGGAGGAGGAGG 1034  
 Db 961 CCGGCTTGTGATGAGTGGGTGTTGGCTCCGATCTGAAAGGAGGAGGAGGAGG 1020  
 QY 1035 CCGGCGGATGAGGAGGAGGAGTCCGCAAGCTCCGTGCTCTTGTGAGCGCACAC 1094  
 Db 1021 CCGGCGGATGAGGAGGAGGAGTCCGCAAGCTCCGTGCTCTTGTGAGCGCACAC 1080  
 QY 1095 CTACCCCAAGGAGGAGGAGCTCGCCCGCCCAAGGAGGAGGAGGAGGAGGAGGAGG 1154  
 Db 1081 CTACCCCAAGGAGGAGGAGGAGCTCGCCCGCCCAAGGAGGAGGAGGAGGAGGAGGAGG 1140

QY 1155 CAGTGGTCCCTCAAGAGACAGAGCCGGGCGAGGCGAGTCCCAAGATTATGAG 1214  
 Db 1141 CAGTGGTCCCTCAAGAGACAGAGCCGGGCGAGGCGAGTCCCAAGATTATGAG 1200  
 QY 1215 GTGGTTTACACTTTTACAGGCCCCCTTGAAGCGGGAAGCTTTCCATTCCTCAACCACTGA 1274  
 Db 1201 GTGGTTTACACTTTTACAGGCCCCCTTGAAGCGGGAAGCTTTCCATTCCTCAACCACTGA 1260  
 QY 1275 ATGTCCCAAGACATGAGTATCCAGACCAAGTCCCGCCCAACAATTTGGGCTCAGGATC 1334  
 Db 1261 ATGTCCCAAGACATGAGTATCCAGACCAAGTCCCGCCCAACAATTTGGGCTCAGGATC 1320  
 QY 1335 TCAGGCTTCCAGTCAAGGAGTACCTTCCGAGCCCTTGTGACTTCCAGGAGTCTGCT 1394  
 Db 1321 TCAGGCTTCCAGTCAAGGAGTACCTTCCGAGCCCTTGTGACTTCCAGGAGTCTGCT 1380  
 QY 1395 CAGTTAGCATTTTCAATGATCATCTTACCCCAAGAGTGGGAGCAACCAAGCAAGC 1454  
 Db 1381 CAGTTAGCATTTTCAATGATCATCTTACCCCAAGAGTGGGAGCAACCAAGCAAGC 1440  
 QY 1455 CCGCTTGAAGATGATCTCTTATGATGAGGTTGCTTATGGGCGCAGGCGAGAGTATG 1514  
 Db 1441 CCGCTTGAAGATGATCTCTTATGATGAGGTTGCTTATGGGCGCAGGCGAGAGTATG 1500  
 QY 1515 CCGCTTGCAGGAGTGAAGAGCAATTCATCACCGAGGAAACCCAGGATTTAAAGAGCC 1574  
 Db 1501 CCGCTTGCAGGAGTGAAGAGCAATTCATCACCGAGGAAACCCAGGATTTAAAGAGCC 1560  
 QY 1575 CTGTGGGCGAGACAGACATAGAGAGGAGTGGAGTCTCCCTTATTCCTGACAAATCTC 1634  
 Db 1561 CTGTGGGCGAGACAGACATAGAGAGGAGTGGAGTCTCCCTTATTCCTGACAAATCTC 1620  
 QY 1635 TACTGATCTTGGCTTTTCTCCGATGCGGATTTGGGGGCGACCTTAAATGCTC 1694  
 Db 1621 TACTGATCTTGGCTTTTCTCCGATGCGGATTTGGGGGCGACCTTAAATGCTC 1680  
 QY 1695 TCTCCAGCCCTGCTCAACCATATCTCAATTTAGTCCAAACCAAGGAGGCTGGACCTTC 1754  
 Db 1681 TCTCCAGCCCTGCTCAACCATATCTCAATTTAGTCCAAACCAAGGAGGCTGGACCTTC 1740  
 QY 1755 CACATATCCATTTGCTTGTGCGCAA 1780  
 Db 1741 CACATATCCATTTGCTTGTGCGCAA 1766

RESULT 2  
 CR608011  
 LOCUS  
 DEFINITION  
 full-length cDNA clone CS0DK010YL01 of Hela cells Cot 25-normalized  
 of Homo sapiens (human).  
 CR608011  
 ACCESSION  
 CR608011.1 GI:50488818  
 VERSION  
 KEYWORDS  
 HTC; CNSLT\_cDNA.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 HOMO SAPIENS  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 1. (bases 1 to 1936)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 JOURNAL  
 REMARK  
 Contact: Feng Liang Email: fliang@lifetech.com URL:  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue  
 2 (bases 1 to 1936)  
 REFERENCE  
 Genoscope.  
 AUTHORS  
 Direct Submission  
 TITLE  
 Submitted (120-JUL-2004) Genoscope - Centre National de Sequencage :  
 JOURNAL  
 BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)  
 Web : www.genoscope.cns.fr  
 COMMENT  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

## FEATURES

RES source	Location/Qualifiers
	1. .1936

1. 1936

/organism="Homo sapiens"

```
/mol_type="mRNA"
```

```
/ab_xref="CAXON:9606"
/c1one="CSNDK010Y1.01"
```

```
/c10ne=c30dKv101Dv1
/c188ve type="Helia ce
```

```
/plasmid="pCMVSPORT 6"
```

```
/plasmid="pCMVSPORT_6
```

## ORIGIN

Query Match	97.48;	Score 1764.4;	DB 3;	Length 1936;
-------------	--------	---------------	-------	--------------

Best Local Similarity	99.9%	Pred. No. 0,
Matched 1765	Concordance 0	Wavelength

Matches 1765; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	4	GTGGGACAGAGGAGACCTCTGCATGTTTAAAGGACCTCCGGGGCCCTCAGAGACGTGGGCGC	103
Dp	1	GTGGGACAGAGGAGACCTCTGCATGTTTAAAGGACCTCTGGGACCTCCTGGGACCTCAGAGACGTGGGCGC	60
QY	106	CGGCCCCGTGGGCGGAGCTCCCCCGCATCCGCGGGGCGCGAATGATCCGGGTGCGCTGCCGAGTG	165
Dp	61	CGGCCCCGTGGGCGGAGCTCCCCCGCATCCGCGGGGCGCGAATGATCCGGGTGCGCTGCCGAGTG	120
QY	166	CTGCTGGGCTGTCTCCCTGGTGTCTGGGTGAAAGTCTGGGTTCTGGGTTTCTGGATTTCG	225
Dp	121	CTGCTGGGCTGTCTCCCTGGTGTCTGGGTGAAAGTCTGGGTTCTGGGTTTCTGGATTTCG	180
QY	226	GGGCGCTTACACGTAGCTGTGACGAGCTCTCGGGGTAGTCCGTCGCGCGCGGCTGCGC	285
Dp	181	GGGCGCTTACACGTAGCTGTGACGAGCTCTCGGGGTAGTCCGTCGCGCGCGGCTGCGC	240
QY	286	CGGGAACGGCCCTTAGCTGTCCGGGGGTCCTCGGGGCCCAAGGCAATTCGGGCTGCAATTGACG	345
Dp	241	CGGGAACGGCCCTTAGCTGTCCGGGGGTCCTCGGGGCCCAAGGCAATTCGGGCTGCAATTGACG	300
QY	346	GGGATCCCGGAGTAGCACCGCGGCGCCCCCGCCCTCACCGAGGGGTCCAGACTGTGTGGGA	405
Dp	301	GGGATCCCGGAGTAGCACCGCGGCGCCCCCGCCCTCACCGAGGGGTCCAGACTGTGTGGGA	360
QY	406	AGNAGGTGCGGGGACGGGTCTCCCTGAGGATCCCGATGCTTACGAGCCCAAGATGCTCAGCTT	465
Dp	361	AGNAGGTGCGGGGACGGGTCTCCCTGAGGATCCCGATGCTTACGAGCCCAAGATGCTCAGCTT	420
QY	466	TATAGGTGTGACCTTACACATGTGACTTCACTCAGTTTGTGATCCTGTAATGGAACAA	525
Dp	421	TATAGGTGTGACCTTACACATGTGACTTCACTCAGTTTGTGATCCTGTAATGGAACAA	480
QY	526	TTGGAAGCTACTTTCACAGGTCTGTGTGAGAGATTAAATGAACATGCTGTAAAGCTT	585
Dp	481	TTGGAAGCTACTTTCACAGGTCTGTGTGAGAGATTAAATGAACATGCTGTGTAAAGCTT	540
QY	586	TTGCGAGAGGAGACCTCGGAGCAGGGCTTGCGCGGACAGCACACTGCTGTACCAAG	645
Dp	541	TTGCGAGAGGAGACCTCGGAGCAGGGCTTGCGCGGACAGCACACTGCTGTACCAAG	600
QY	646	GACCAAGGACACATGAAAGACCCCGGTGAGAGCTGCGCTCAGTGGATGACAGACCTTCGG	705
Dp	601	GACCAAGGACACATGAAAGACCCCGGTGAGAGCTGCGCTCAGTGGATGACAGACCTTCGG	660
QY	706	CCTTCAGCACCGCTGCCAGGTGTGCTACCGGGTCMAAGCCAGAGCGTCATATGTGATGA	765
Dp	661	CCTTCAGCACCGCTGCCAGGTGTGCTACCGGGTCMAAGCCAGAGCGTCATATGTGATGA	720
QY	766	GACTCTGTTTGGGACGCCAGCAGGACCTCGGCGTACCTCAACGGACTTTCGATCCGCCCTG	825
Dp	721	GACTCTGTTTGGGACGCCAGCAGGACCTCGGCGTACCTCAACGGACTTTCGATCCGCCCTG	780
QY	826	GGTGAAGAAAGCTTACAGAAACAGAGGCGTGTGGGCAAGAGGCAATCGAAGGCTTTGGGGGC	885
Dp	781	GGTGAAGAAAGCTTACAGAAACAGAGGCGTGTGGGCAAGAGGCAATCGAAGGCTTTGGGGGC	840
QY	886	AAAGGGAGCTGTGAGACCACTCCCTCAAGGGGACGACCTCCCACTTCAACCAAGAA	945

Db	841	AAAGGGAGCTGTGAGACCAACCCCTTAAAGGGAGGACACCCCACTCTCAGACCAAGAA	900
OY	946	GAAGAACAAATACAGACCCCATCAGCCACACCCCGCTTATCTGTATGATGCTGTTTGG	1005
Db	901	GAAGAACAAATACAGACCCCATCAGCCACACCCCGCTTATCTGTATGATGCTGTTTGG	960
OY	1006	CTCCCGATCTGTAAGGGCCGAGCTTCGGGGGCCCCGGATGGCGAAGGGGGATGCGCAAA	10655
Db	961	CTCCCGATCTGTAAGGGCCGAGCTTCGGGGGCCCCGGATGGCGAAGGGGGATGCGCAAA	1020
OY	1066	GCTCCGTCGCTCTCTTGTGACGCCACCACTAACCCCGAGGGGTAGCACTGCGCCGCC	1125
Db	1021	GCTCCGTCGCTCTCTTGTGACGCCACCACTAACCCCGAGGGGTAGCACTGCGCCGCC	1080
OY	1126	CAGGAGGACCACTGCGAGCATTTACCCAGCTGTCTCTTCAAGACAGAGCCGGGCGC	1185
Db	1081	CAGGAGGACCACTGCGAGCATTTACCCAGCTGTCTCTTCAAGACAGAGCCGGGCGC	1140
OY	1186	AGCGGACAGCTCCCAAGATATCTATAGGGTGGGTTTACCTTTACGGCCCCCGAAGGC	1245
Db	1141	AGCGGACAGCTCCCAAGATATCTATAGGGTGGGTTTACCTTTACGGCCCCCGAAGGC	1200
OY	1246	GGAATCTTCCATTCCTCTCAACCACTGAATGTGCCAGCACTGTCATCCAGCCACAG	1305
Db	1201	GGAATCTTCCATTCCTCTCAACCACTGAATGTGCCAGCACTGTCATCCAGCCACAG	1260
OY	1306	TGCCCCCACAACAATGGGCTCAGAGATCTCAGGCTTTCCAGTCAGGGGTGACCTTCCG	1365
Db	1261	TGCCCCCACAACAATGGGCTCAGAGATCTCAGGCTTTCCAGTCAGGGGTGACCTTCCG	1320
OY	1366	GAGCCCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCATTTCACTGTCATTCACCAAG	1425
Db	1321	GAGCCCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCATTTCACTGTCATTCACCAAG	1380
OY	1426	ACGAGTGGGGCCACCCAGAAACCAAAAGCCCTCTGGAAATGATATCTTTTCATCAGGGT	1485
Db	1381	ACGAGTGGGGCCACCCAGAAACCAAAAGCCCTCTGGAAATGATATCTTTTCATCAGGGT	1440
OY	1486	TGCTTATGGGGCCACGGCGACAGTATAGGCCCTTCGCAAGGATGAGAGCATTCATCAC	1545
Db	1441	TGCTTATGGGGCCACGGCGACAGTATAGGCCCTTCGCAAGGATGAGAGCATTCATCAC	1500
OY	1546	CCAAGGAAACCCAGAGTATTAAGAAGGCCCTGTGGGGGACAGACATAGCAGGGGTGG	1605
Db	1501	CCAAGGAAACCCAGAGTATTAAGAAGGCCCTGTGGGGGACAGACATAGCAGGGGTGG	1560
OY	1606	GCACTGCTCTCTTATCTGACATCTCTAGTGCATTTCTTGCTTTTCTCCGATTCG	1665
Db	1561	GCACTGCTCTCTTATCTGACATCTCTAGTGCATTTCTTGCTTTTCTCCGATTCG	1620
OY	1666	GGATTTTGGGGGCCACCTCTAAGATGCTCTCTCAGGCCCTGTCTCAACCAATCTCAAAAT	1725
Db	1621	GGATTTTGGGGGCCACCTCTAAGATGCTCTCTCAGGCCCTGTCTCAACCAATCTCAAAAT	1680
OY	1726	TAGTGCCAACCCAGGGGCTTGCACTTCCACATCATCTATGCTTGTCTGCCAAGTGG	1785
Db	1681	TAGTGCCAACCCAGGGGCTTGCACTTCCACATCATCTATGCTTGTCTGCCAAGTGG	1740
OY	1786	AATTAACGGGTGATTTGCCAACTGG 1811	
Db	1741	AATTAACGGGTGATTTGCCAACTGG 1766	

RESULT 3				
LOCUS	CR622111	1839 bp	mRNA	linear HTc 21-JUL-2004
DEFINITION	full-length cDNA clone CS0DA004YD23 of Neuroblastoma of Homo sapiens (human) .			
ACCESSION	CR622111			
VERSION	CR622111.1	GI:50502918		
KEYWORDS	HTC; CNSLT_cDNA.			
SOURCE	Homo sapiens (human)			

REFERENCE	AUTHORS	JOURNAL	TITLE
1	(bases 1 to 1839)		
2	(bases 1 to 1839)		
3	(bases 1 to 1839)		
4	(bases 1 to 1839)		
5	(bases 1 to 1839)		
6	(bases 1 to 1839)		
7	(bases 1 to 1839)		
8	(bases 1 to 1839)		
9	(bases 1 to 1839)		
10	(bases 1 to 1839)		
11	(bases 1 to 1839)		
12	(bases 1 to 1839)		
13	(bases 1 to 1839)		
14	(bases 1 to 1839)		
15	(bases 1 to 1839)		
16	(bases 1 to 1839)		
17	(bases 1 to 1839)		
18	(bases 1 to 1839)		
19	(bases 1 to 1839)		
20	(bases 1 to 1839)		
21	(bases 1 to 1839)		
22	(bases 1 to 1839)		
23	(bases 1 to 1839)		
24	(bases 1 to 1839)		
25	(bases 1 to 1839)		
26	(bases 1 to 1839)		
27	(bases 1 to 1839)		
28	(bases 1 to 1839)		
29	(bases 1 to 1839)		
30	(bases 1 to 1839)		
31	(bases 1 to 1839)		
32	(bases 1 to 1839)		
33	(bases 1 to 1839)		
34	(bases 1 to 1839)		
35	(bases 1 to 1839)		
36	(bases 1 to 1839)		
37	(bases 1 to 1839)		
38	(bases 1 to 1839)		
39	(bases 1 to 1839)		
40	(bases 1 to 1839)		
41	(bases 1 to 1839)		
42	(bases 1 to 1839)		
43	(bases 1 to 1839)		
44	(bases 1 to 1839)		
45	(bases 1 to 1839)		
46	(bases 1 to 1839)		
47	(bases 1 to 1839)		
48	(bases 1 to 1839)		
49	(bases 1 to 1839)		
50	(bases 1 to 1839)		
51	(bases 1 to 1839)		
52	(bases 1 to 1839)		
53	(bases 1 to 1839)		
54	(bases 1 to 1839)		
55	(bases 1 to 1839)		
56	(bases 1 to 1839)		
57	(bases 1 to 1839)		
58	(bases 1 to 1839)		
59	(bases 1 to 1839)		
60	(bases 1 to 1839)		
61	(bases 1 to 1839)		
62	(bases 1 to 1839)		
63	(bases 1 to 1839)		
64	(bases 1 to 1839)		
65	(bases 1 to 1839)		
66	(bases 1 to 1839)		
67	(bases 1 to 1839)		
68	(bases 1 to 1839)		
69	(bases 1 to 1839)		
70	(bases 1 to 1839)		
71	(bases 1 to 1839)		
72	(bases 1 to 1839)		
73	(bases 1 to 1839)		
74	(bases 1 to 1839)		
75	(bases 1 to 1839)		
76	(bases 1 to 1839)		
77	(bases 1 to 1839)		
78	(bases 1 to 1839)		
79	(bases 1 to 1839)		
80	(bases 1 to 1839)		
81	(bases 1 to 1839)		
82	(bases 1 to 1839)		
83	(bases 1 to 1839)		
84	(bases 1 to 1839)		
85	(bases 1 to 1839)		
86	(bases 1 to 1839)		
87	(bases 1 to 1839)		
88	(bases 1 to 1839)		
89	(bases 1 to 1839)		
90	(bases 1 to 1839)		
91	(bases 1 to 1839)		
92	(bases 1 to 1839)		
93	(bases 1 to 1839)		
94	(bases 1 to 1839)		
95	(bases 1 to 1839)		
96	(bases 1 to 1839)		
97	(bases 1 to 1839)		
98	(bases 1 to 1839)		
99	(bases 1 to 1839)		
100	(bases 1 to 1839)		

Db	542	TCGACAGCCCTCGGCGCTTCAGCACCCGCTGCCAGGTGGCTACCGGGTCAAGGCGAGACGT	601
Oy	753	CATATGTGGATTAAGA CTCTGTTTGGCAGCCCGACAGGAGACCCGGGCTTACCCCA CGGACT	812
Db	602	CATATGTGGATTAAGA CTCTGTTTGGCAGCCCGACAGGAGACCCGGGCTTACCCCA CGGACT	661
Oy	813	TCGATCCGCGCTTGGGTGGAGAAAGGCTTAA CAGAACCAAGGCGGTGGGCAAGAGGATCGA	872
Db	662	TCGATCCGCGCTTGGGTGGAGAAAGGCTTAA CAGAACCAAGGCGGTGGGCAAGAGGATCGA	721
Oy	873	AGGCGTTTGGGGGCAAAAGGGAGCGTGTAGACACACCCCTTCAAGGGGGGCGACCCCGACCC	932
Db	722	AGGCGTTTGGGGGCAAAAGGGAGCGTGTAGACACACCCCTTCAAGGGGGGCGACCCCGACCC	781
Oy	933	TCACACCAAGGAAGAAAGAAATTCACAGCCCATCAGCCACACCCCGTCTTA CTGTGATG	992
Db	782	TCACACCAAGGAAGAAAGAAATTCACAGCCCATCAGCCACACCCCGTCTTA CTGTGATG	841
Oy	993	AGTCGCTGTTTGGCTCCCGATCTGAAGGCGC CAGCTTGGGGGCCCGCGGATGGCGAAGG	1053
Db	842	AGTCGCTGTTTGGCTCCCGATCTGAAGGCGC CAGCTTGGGGGCCCGCGGATGGCGAAGG	901
Oy	1053	GGGAGCCGGGCAAAAGTCGCGTCTCTTGTGTAGACGCGCACCTTACCCCGAGGGGTATACC	1112
Db	902	GGGAGCCGGGCAAAAGTCGCGTCTCTTGTGTAGACGCGCACCTTACCCCGAGGGGTATACC	961
Oy	1113	ACTCGCCCCGCGCCCGACGAGGAGCACCA CTGCGAGCCATTACCCAGCTGTCTCTTCAAGA	1172
Db	962	ACTCGCCCCGCGCCCGACGAGGAGCACCA CTGCGAGCCATTACCCAGCTGTCTCTTCAAGA	1022
Oy	1173	CAGAGCCGGGGGCGACGGGCGAGCTTCCGAGAA GTTATCTATGSGTGGTTTACA CTCTTAC	1232
Db	1022	CAGAGCCGGGGGCGACGGGCGAGCTTCCGAGAA GTTATCTATGSGTGGTTTACA CTCTTAC	1081
Oy	1233	GCCCCGTTAAACGGGGGACTTTTCCATTCCCTTCCACCCACTGAATGTCCCGGACGACTGGTC	1292
Db	1082	GCCCCGTTAAACGGGGGACTTTTCCATTCCCTTCCACCCACTGAATGTCCCGGACGACTGGTC	1141
Oy	1293	ATCCAGCCACCAAGTGC CCCCCCA CCAATATGGAGCTCAGAGTCTCAGAGGCTTCCAGCTGAG	1352
Db	1142	ATCCAGCCACCAAGTGC CCCCCCA CCAATATGGAGCTCAGAGTCTCAGAGGCTTCCAGCTGAG	1201
Oy	1353	GGGTGA CTTTCGGAGCCCTCTGTGTGA CTTCAGGGGCTCGCTCAGTTTGA CATTTCA GTGC	1412
Db	1202	GGGTGA CTTTCGGAGCCCTCTGTGTGA CTTCAGGGGCTCGCTCAGTTTGA CATTTCA GTGC	1261
Oy	1413	CATCTAACCCCA CAGAGGTGGGGCCACCCGAGAA ACCCAAACCCCTTGGAAATGATACT	1472
Db	1262	CATCTAACCCCA CAGAGGTGGGGCCACCCGAGAA ACCCAAACCCCTTGGAAATGATACT	1321
Oy	1473	CTTTTCATCAGAGGTGTGCCTATGATGGGCG CACAGGATATGGCCCTTGGCAGGGGTAGGA	1532
Db	1322	CTTTTCATCAGAGGTGTGCCTATGATGGGCG CACAGGATATGGCCCTTGGCAGGGGTAGGA	1381
Oy	1533	GGA CATTTCATCACCCAGGGAACCC CAGGTATTTAAAGAA GCCCTGTGGGGGCGACAGAC	1592
Db	1382	GGA CATTTCATCACCCAGGGAACCC CAGGTATTTAAAGAA GCCCTGTGGGGGCGACAGAC	1441
Oy	1593	ATAGCAGAGGGGTGGGCGACGTGCTCCCTTTATCTGTGACAATCTCTAGTGA TTTTGGCTTT	1652
Db	1442	ATAGCAGAGGGGTGGGCGACGTGCTCCCTTTATCTGTGACAATCTCTAGTGA TTTTGGCTTT	1501
Oy	1653	TTTCTCCCGA TTTGGGAGGCGACCTCTAAGATGGCCTCTCTCCAGGCCCTGTCTCAA	1712
Db	1502	TTTCTCCCGA TTTGGGAGGCGACCTCTCTAAGATGGCCTCTCTCTCCAGGCCCTGTCTCAA	1561
Oy	1713	CCATATCTCAAAATTAAGTGC CAAACCCAGAGGGGCTTGGCACTTCCCA CATATCATATGTCTTT	1772
Db	1562	CCATATCTCAAAATTAAGTGC CAAACCCAGAGGGGCTTGGCACTTCCCA CATATCATATGTCTTT	1621
Oy	1773	GCTGCGAAATGCGAATTAACGGCGTGA TTTGCGAACTGG	1811
Db	1622	GCTGCGAAATGCGAATTAACGGCGTGA TTTGCGAACTGG	1660

[illegible]

OY	573	CTTGAAAGCTCTTTGACAGAGAGGAGCTCCGGAACAAGGAGCTTGCGCCGACAGACACA	632
Db	422	CTTGAAAGCTCTTTGACAGAGAGGAGCTCCGGAACAAGGAGCTTGCGCCGACAGACACA	481
OY	633	TGCTGTACAAGGACCAAGGAGCATGAAAGACCCCGTGGAGCTGCGCTCATGGGA	692
Db	482	TGCTGTACAAGGAGACCAAGGAGCATGAAAGACCCCGTGGAGCTGCGCTCATGGGA	541
OY	693	TGCAAGCCCTGGGACCTTAGCAACGCTCCGAGGTGGCTACCGGGGTCAAGGACGGAAGT	752
Db	542	TGCAAGCCCTGGGACCTTAGCAACGCTCCGAGGTGGCTACCGGGGTCAAGGACGGAAGT	601
OY	753	CATATGTGATGAGACTCTGTTTGGACAGCCCGACAGGAGACCCCGGCTTACCCACCGGACT	812
Db	602	CATATGTGATGAGACTCTGTTTGGACAGCCCGACAGGAGACCCCGGCTTACCCACCGGACT	661
OY	813	TCGATTCGCCCTCTGGGTGGAGAAAGGTAAACAACAAGAGGCTGGGACAGAGGACATCGA	872
Db	662	TCGATTCGCCCTCTGGGTGGAGAAAGGTAAACAACAAGAGGCTGGGACAGAGGACATCGA	721
OY	873	AGGCTTTGGGGACAAAGGGAGCTGTGAGACACCCCTCAAGGGGACAGCACCCACCC	932
Db	722	AGGCTTTGGGGACAAAGGGAGCTGTGAGACACCCCTCAAGGGGACAGCACCCACCC	781
OY	933	TCACACCAAGAGAGAAACAATACAGACCCTATGAGCAACCCGCTTACTGTGATG	992
Db	782	TCACACCAAGAGAGAAACAATACAGACCCTATGAGCAACCCGCTTACTGTGATG	841
OY	993	AGTGGCTGTGGCTCCCGACTGTAGAAAGGCGCAGCTTGAGGGGCGCCGGGAGAGGAGAG	105
Db	842	AGTGGCTGTGGCTCCCGACTGTAGAAAGGCGCAGCTTGAGGGGCGCCGGGAGAGGAGAG	901
OY	1053	GGAGTGCAGAAAGCTCCGCTCTCTTGTGAGCGCCACACTTACCCGAGGGTAGGC	1111
Db	902	GGAGTGCAGAAAGCTCCGCTCTCTTGTGAGCGCCACACTTACCCGAGGGTAGGC	961
OY	1113	ACTGGCCCCGCCCCAGGGAGGACACATGCGAGGCTTACCCAGCTGTCCTTCCAGA	1177
Db	962	ACTGGCCCCGCCCCAGGGAGGACACATGCGAGGCTTACCCAGCTGTCCTTCCAGA	1022
OY	1173	CAGAGCGGGGCGAGCGGACAGTCCCGAAGTTATCTATGGGTGGTTACACTCTTCAC	123
Db	1022	CAGAGCGGGGCGAGCGGACAGTCCCGAAGTTATCTATGGGTGGTTACACTCTTCAC	108
OY	1233	GCCCCCTGAAACGGGGACCTTCCCATTCCTCAACCCACTGATATGTCGCCAGACTGTC	129
Db	1082	GCCCCCTGAAACGGGGACCTTCCCATTCCTCAACCCACTGATATGTCGCCAGACTGTC	114
OY	1293	ATTCAGGCAACAGTGGCCCCCAACAATGGGCGTCAAGAGATCTGAGGCTTCCAGTCA	135
Db	1142	ATTCAGGCAACAGTGGCCCCCAACAATGGGCGTCAAGAGATCTGAGGCTTCCAGTCA	120
OY	1353	GGGTGACCTTCCGAGAGCCCTGTGTGACTTCAAGGCTCGCTCAGTTAGCAATTCAGTGC	141
Db	1202	GGGTGACCTTCCGAGAGCCCTGTGTGACTTCAAGGCTCGCTCAGTTAGCAATTCAGTGC	126
OY	1413	CATCTAACCCCAAGAGGTGGGCGACCCAGAAACCAAAGCCCTTGGAAATGATACT	147
Db	1262	CATCTAACCCCAAGAGGTGGGCGACCCAGAAACCAAAGCCCTTGGAAATGATACT	132
OY	1473	CTTTCATCAGAGGTGCTATGAGGCGACAGGAGATGGGCCCTTGCCAGGGTAGGA	153
Db	1322	CTTTCATCAGAGGTGCTATGAGGCGACAGGAGATGGGCCCTTGCCAGGGTAGGA	138
OY	1533	GGAATTATCATCACCCAGGGAAACCCAGAGTATTAAAGAAAGCCCTGTGGGGCAAGACAG	159
Db	1382	GGAATTATCATCACCCAGGGAAACCCAGAGTATTAAAGAAAGCCCTGTGGGGCAAGACAG	144
OY	1593	ATTAGCAGGGGTGGGCAATGCTCCCTTATTCGTACAAATCTATAGTGAATCTTGCCCTT	165
Db	1442	ATTAGCAGGGGTGGGCAATGCTCCCTTATTCGTACAAATCTATAGTGAATCTTGCCCTT	150

QY	1653	TTCTCCCGAGTTGGGGGGGACCTCTAAGAGCCCTCTCCAGCCCTCTCA	1712
Db	1502	TTCTCCCGAGTTGGGGGGGACCTCTAAGATCCCTCTCCAGCCCTCTCA	1551
QY	1713	CCATACCTCAAAATTAGTGCACACCGAGGGGCTGCACCTCCACATCATCTT	1772
Db	1562	CCATACCTCAAAATTAGTGCACACCGAGGGGCTGCACCTCCACATCATCTT	1621
QY	1773	GCTGCCAAGT 1783	
Db	1622	GCTGCCAAGT 1632	
RESULT 5			
LOCUS	CR621332	1614 bp	mRNA
DEFINITION	Full-length cDNA clone CS0DJ008Y14 of T cells (Jurkat cell line)		
ACCESSION	CR621332		
VERSION	CR621332.1	GI:50502139	
KEYWORDS	HTC; CNSLT cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Li, W.B., Gruber, C., Jesssee, J., and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished		
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 1614) Genoscope.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr		
TITLE	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr		
JOURNAL	- Web : www.genoscope.cns.fr)		
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.		
FEATURES	Location/Qualifiers		
source	1..1614		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DJ008Y14"		
	/tissue_type="T cells (Jurkat cell line) Cot		
	10-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
	Query Match 89.1%; Score 1614; DB 3; Length 1614;		
	Best Local Similarity 100.0%; Pred. No. 0;		
	Matches 1614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	160	GCAAGTCTGCTGAGCTGCTCCCTGCTGCTGAGGTGCAAGTCTGGGTTTCTGG	219
Db	1	GCAAGTCTGCTGAGCTGCTCCCTGCTGCTGAGGTGCAAGTCTGGGTTTCTGG	60
QY	220	ATTGCGGGGCGGTTTACACCGTAAGCTGTGCGCGGCTCTCGGGTAAATCCGTCGGCGGCG	279
Db	61	ATTGCGGGGCGGTTTACACCGTAAGCTGTGCGCGGCTCTCGGGTAAATCCGTCGGCGGCG	120
QY	280	GTGCCCCGAGAGCGCTAGCTGCGCGGGAGTCCGAGGCCACAGCATTTCCGGCTGACAGA	339
Db	121	GTGCCCCGAGAGCGCTAGCTGCGCGGGAGTCCGAGGCCACAGCATTTCCGGCTGACAGA	180
QY	340	TTGAGCGGGAGATCCGGAGATCACCGCGGCGCCCGCGCCCTTACCGACGGAGTCCAGACTG	399
Db	181	TTGAGCGGGAGATCCGGAGATCACCGCGGCGCCCGCGCCCTTACCGACGGAGTCCAGACTG	240

OY	400	GTGGAGAGAGTGTCCGGGAGCGGGTCCCTGAGATCCCGATGCTTACGAGCCAAAGATGCT	459
Db	241	GTGGGAGAGAGTGTCCGGGAGCGGGTCCCTGAGATCCCGATGCTTACGAGCCAAAGATGCT	300
OY	460	CAGCTTTATAGAGTGTGACCTACATGTGACTTCACGTCAGTTTGTGATCCCGTAAATG	519
Db	301	CAGCTTTATAGTGTGACCTACATGTGACTTCACGTTTGTGATCCCGTAAATG	360
OY	520	GACAAATTCAGAGCTACTTCACTAGTGTGTGAGAGATTAAATGAAACAATGCTTTAA	579
Db	361	GACAAATTCAGAGCTACTTCACTAGTGTGTGAGAGATTAAATGAAACAATGCTTTAA	420
OY	580	AGCTCTTTGCAAGAGGGAGCTCGGAAGCAAGGCTTGGCCGGGAGAGCAACCTGCTGTG	639
Db	421	AGCTCTTTGCAAGAGGGAGCTCGGAAGCAAGGCTTGGCCGGGAGAGCAACCTGCTGTG	480
OY	640	ACCAAGGAGCAACAGGACAGATGAAAGACCCCGGTGAGCTGGCCGTCACTAGTGGATGACAG	699
Db	481	ACCAAGGAGCAACAGGACAGATGAAAGACCCCGGTGAGCTGGCCGTCACTAGTGGATGACAG	540
OY	700	CCTCGGCTTTGAGACACCGCTGCGAGGTGTGCTACCGGGTCAAGGCGAGAGCTCATATGT	759
Db	541	CCTCGGCTTTGAGACACCGCTGCGAGGTGTGCTACCGGGTCAAGGCGAGAGCTCATATGT	600
OY	760	GGATGAGACTGTGTTTGGCAGCCCAAGAGGCAACCCGCGCTACCCACCGGACTTTCATATCC	819
Db	601	GGATGAGACTGTGTTTGGCAGCCCAAGAGGCAACCCGCGCTACCCACCGGACTTTCATATCC	660
OY	820	GCCTTGAGTGTGAGAGGCTTAAACAGAACAGAGGCGTGGGCAAGAGGCTATGAAAGGCTT	879
Db	661	GCCTTGAGTGTGAGAGGCTTAAACAGAGGCGTGGGCAAGAGGCTATGAAAGGCTT	720
OY	880	GGGGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAGGGGGCAACAACCCCTCACACC	939
Db	721	GGGGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAGGGGGCAACAACCCCTCACACC	780
OY	940	AAGGAAGAAGAACAAATACAGACCCCATACAGCAACCCGCTTTATCTGTATGATGTGCT	999
Db	781	AAGGAAGAAGAACAAATACAGACCCCATACAGCAACCCGCTTTATCTGTATGATGTGCT	840
OY	1000	GTTTGTGCTCCGATCTGAAGGCGCAGCTTCCGGGGCCCGCGGATGTGCGAAGGGAGATGC	1059
Db	841	GTTTGTGCTCCGATCTGAAGGCGCAGCTTCCGGGGCCCGCGGATGTGCGAAGGGAGATGC	900
OY	1060	CGAAAGCTCCGTGCTCTTTGTGAGCGCAACAACCTACCCCAAGGGGTAGCCACTCGCC	1119
Db	901	CGAAAGCTCCGTGCTCTTTGTGAGCGCAACAACCTACCCCAAGGGGTAGCCACTCGCC	960
OY	1120	CCGGCCCAAGGAGGAGCAACATGCGAGGCAATTACCCGAGCTGTGCTCCCTCAAGACAGAGCC	1179
Db	961	CCGGCCCAAGGAGGAGCAACATGCGAGGCAATTACCCGAGCTGTGCTCCCTCAAGACAGAGCC	1020
OY	1180	GGGGCCAGCGGACAGCTCCAGAGATTATCTATGGTGGGTACAATTCTTCAACGCCCCCT	1239
Db	1021	GGGGCCAGCGGAGAGCTCCAGAGATTATCTATGGTGGGTACAATTCTTCAACGCCCCCT	1080
OY	1240	GAAAGCGGGGACTTTCCATTCCTTCAACCCACTGGAATGTCCCAAGCACTGTGATCCAGC	1299
Db	1081	GAAAGCGGGGACTTTCCATTCCTTCAACCCACTGGAATGTCCCAAGCACTGTGATCCAGC	1140
OY	1300	CACCAAGTCCCCCAACAATGGGCTTCAAGGATCTTAGGCTTTCAACGTCAGGGGTGAC	1359
Db	1141	CACCAAGTCCCCCAACAATGGGCTTCAAGGATCTTAGGCTTTCAACGTCAGGGGTGAC	1200
OY	1360	CTTCCGAGGCCCTGTGTGACTTCCAGAGGCTGTGCTAGTAGAATTTCACTGTGCATCTAC	1419
Db	1201	CTTCCGAGGCCCTGTGTGACTTCCAGAGGCTGTGCTAGTAGAATTTCACTGTGCATCTAC	1260
OY	1420	CCCAAGAGAGTGTGGGCGCAACCAAGAACCAAAAGCCCTTTGGAATGATATCTTTGAT	1479
Db	1261	CCCAAGAGAGTGTGGGCGCAACCAAGAACCAAAAGCCCTTTGGAATGATATCTTTGAT	1320



QY 1480 CAGGTTGCTTATGGGGCCACGCGACAGATATGCCCCCTTGCACAGGATAGAGACATT 1539  
DB 1331 CAGGTTGCTTATGGGGCCACGCGACAGATATGCCCCCTTGCACAGGATAGAGACATT 1380  
QY 1540 CATCACCAAGGAAACCCAGATTAAGAGAGCCCTGTGGGGGACAGACATATAGCAG 1599  
DB 1381 CATCACCAAGGAAACCCAGATTAAGAGAGCCCTGTGGGGGACAGACATATAGCAG 1440  
QY 1600 GGGTGGGAGTGGCCCTCTTATTCCTGACAAATCTCTAGTCGATTCCTGCTTTTCTCCC 1659  
DB 1441 GGGTGGGAGTGGCCCTCTTATTCCTGACAAATCTCTAGTCGATTCCTGCTTTTCTCCC 1500  
QY 1660 GATTGCGATTTGGGGGACACCTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCACTACT 1719  
DB 1501 GATTGCGATTTGGGGGACACCTCTAAGATGCTCTCTCTCCAGCCCTGTCTCAACCACTACT 1560  
QY 1720 CCAATTAATGTCGAACCCAGAGGGGCTGGGACCTCCCAATCATTCATTGCTTGG 1773  
DB 1561 CCAATTAATGTCGAACCCAGAGGGGCTGGGACCTCCCAATCATTCATTGCTTGG 1614

RESULT 6  
CR600183  
LOCUS CR600183 1780 bp mRNA linear HTC 21-JUL-2004  
DEFINITION Full-length cDNA clone CS0DC011YN03 of Neuroblastoma Cot  
ACCESSION CR600183  
VERSION CR600183.1 GI:50480990  
KEYWORDS HTC; NSLT\_cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Peng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paraday Avenue  
2 (bases 1 to 1780)  
Genoscope.  
Direct Submission  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
Location/Qualifiers  
Source 1..1780  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC011YN03"  
/tissue\_type="Neuroblastoma Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 88.3%; Score 1600; DB 3; Length 1780;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 GTTCTGATTCGCGGGCCGTTCAACGATAGCTGTGCGGGTCTCTGGGTAGTCCGTC 271  
DB 1 GTTCTGATTCGCGGGCCGTTCAACGATAGCTGTGCGGGTCTCTGGGTAGTCCGTC 60  
QY 272 GCGCGCGGTGCGCGGAGCGGCTTACGCTGCGGGGCTCGGGGCCCAAGGCAATTCGG 331  
DB 61 GCGCGCGGTGCGCGGAGCGGCTTACGCTGCGGGGCTCGGGGCCCAAGGCAATTCGG 120

QY 332 GCTGACGATTTGACGGGGATCCCGAGATGACACGCGGCCCCCGGCTCTACCGAGGGTTC 391  
DB 121 GCTGACGATTTGACGGGGATCCCGAGATGACACGCGGCCCCCGGCTCTACCGAGGGTTC 180  
QY 392 CAGACCTGTGGGAAAGAAAGTTCGGGGAGCGGGTCCCTGAGATCCCGATGCTTACAGGC 451  
DB 181 CAGACCTGTGGGAAAGAAAGTTCGGGGAGCGGGTCCCTGAGATCCCGATGCTTACAGGC 240  
QY 452 AAGATGCTCAGCTTTATAGGTGTGACATACATATGTACTTCACTCACTTTTGTGATCC 511  
DB 241 AAGATGCTCAGCTTTATAGGTGTGACATACATATGTACTTCACTCACTTTTGTGATCC 300  
QY 512 GTAATATGACCAATTTGAGATCTTCAAGTGTGTTGAGAGATTAATGAAACAAT 571  
DB 301 GTAATATGACCAATTTGAGATCTTCAAGTGTGTTGAGAGATTAATGAAACAAT 360  
QY 572 GCTTGAAGCTCTTTGACGAGGAGAGGCTCTGGAAGAGGGCTGGCCGCGACAGACAC 631  
DB 361 GCTTGAAGCTCTTTGACGAGGAGAGGCTCTGGAAGAGGGCTGGCCGCGACAGACAC 420  
QY 632 CTGCTGTCAACGAGGACCAAGGACGATGAAGACCCCGGTGAGCTGCTCACTGCG 691  
DB 421 CTGCTGTCAACGAGGACCAAGGACGATGAAGACCCCGGTGAGCTGCTCACTGCG 480  
QY 692 ATGCAAGCCCTGAGCTTACAGACCGCTGCGAGGAGTACCGGGGTCAAGGCGCAGAGAG 751  
DB 481 ATGCAAGCCCTGAGCTTACAGACCGCTGCGAGGAGTACCGGGGTCAAGGCGCAGAGAG 540  
QY 752 TCATATGTGATGATGACTCTGTTTGGCAAGCCGAGAGGACCCGACTTACCCACGAGAC 811  
DB 541 TCATATGTGATGATGACTCTGTTTGGCAAGCCGAGAGGACCCGACTTACCCACGAGAC 600  
QY 812 TTGCAATCCGCTCTGGGTGAGAAAGCTTAACAGAACGAGGTGTGGCAAGAGGATGCG 871  
DB 601 TTGCAATCCGCTCTGGGTGAGAAAGCTTAACAGAACGAGGTGTGGCAAGAGGATGCG 660  
QY 872 AAGGCTTGGGGGGCAAGGGGAGCTGTGAGACACCCCTCAAGGGGCGACGCCCAACC 931  
DB 661 AAGGCTTGGGGGGCAAGGGGAGCTGTGAGACACCCCTCAAGGGGCGACGCCCAACC 720  
QY 932 CTCACACCAAGGAAAGAAACAATATACAGACCCATCAGACACCCCGTCTTACTGTGAT 991  
DB 721 CTCACACCAAGGAAAGAAACAATATACAGACCCATCAGACACCCCGTCTTACTGTGAT 780  
QY 992 GAGTGTGCTTTGCTCTCCGATCTGAAGCGCGACCTTGGGGGCCCGGAGTGGGAG 1051  
DB 781 GAGTGTGCTTTGCTCTCCGATCTGAAGCGCGACCTTGGGGGCCCGGAGTGGGAG 840  
QY 1052 GGGGATGCGGAAAGCTCGGTCTCTTGTGAGAGCCACACCTAACCCCGAGGGGTGCG 1111  
DB 841 GGGGATGCGGAAAGCTCGGTCTCTTGTGAGAGCCACACCTAACCCCGAGGGGTGCG 900  
QY 1112 CACTGCCCCCGCCAGAGGAGGACACACTGCGAGGCAATTCACCGAGTGTCTCTCAAG 1171  
DB 901 CACTGCCCCCGCCAGAGGAGGACACACTGCGAGGCAATTCACCGAGTGTCTCTCAAG 960  
QY 1172 ACAAGCGCGGGGCGACGCGAGACTTCCAGAAATTTATCTATGAGGTGTTACACTCTTCA 1231  
DB 961 ACAAGCGCGGGGCGACGCGAGACTTCCAGAAATTTATCTATGAGGTGTTACACTCTTCA 1020  
QY 1232 GGGCCCGTGAAGCGGGGACTTTCCATTCCTCTCAACCCCTGAAGTGTCCCGAGCACTGGT 1291  
DB 1021 GGGCCCGTGAAGCGGGGACTTTCCATTCCTCTCAACCCCTGAAGTGTCCCGAGCACTGGT 1080  
QY 1292 CATCAGGCAACGATGCCCCCAACAATATGAGGCTCTGAGATCTGAGGCTTTCACAGTCA 1351  
DB 1081 CATCAGGCAACGATGCCCCCAACAATATGAGGCTCTGAGATCTGAGGCTTTCACAGTCA 1140  
QY 1352 GGGGTGACCTTTCGAGGAGCCCTGTGTGACTTTCAGAGGCTCTGCTCACTTTCAGT 1411  
DB 1141 GGGGTGACCTTTCGAGGAGCCCTGTGTGACTTTCAGAGGCTCTGCTCACTTTCAGT 1200  
QY 1412 CCATCTACCCACGACGAGGTGGGGCCACCCAGAAACCAAGGCCCTTGGAAATGATAC 1471

Db	1201	CCATTTA	CCCCCAGCAGAGGTGGGGCCACCAGAAACCAAGCCCCCTTGGAAATGATAC	1260
OY	1472	TCCTTCATC	AGGGTTCCTATGGGGCCACGGCGAGGTATGGCCCTTGGCCAGGTAGG	1531
Db	1261	TCCTTCATC	AGGGTTCCTATGGGGCCACGGCGAGGTATGGCCCTTGGCCAGGGTAGG	1320
OY	1532	AGGACATTC	ATCACCAGAGAAACCCAGGTATTAAAGAAAGCCCTGTGGGGGGCAGACAGA	1593
Db	1331	AGGACATTC	ATCACCAGAGAAACCCAGGTATTAAAGAAAGCCCTGTGGGGGGCAGACAGA	1380
OY	1592	CATACGAG	GGGTGGGCACTGCCCTTATCTCTACATCTCTAAGTCATTTCTTGCCCTT	1651
Db	1381	CATACGAG	GGGTGGGCACTGCCCTTATCTCTACATCTCTAAGTCATTTCTTGCCCTT	1440
OY	1652	TTTCTCCG	ATTCGGGATTTGGGGGCGACCTTGAAGTGCCTCTTCCAGGCCCTGTCTCA	1711
Db	1441	TTTCTCCG	ATTCGGGATTTGGGGGCGACCTTGAAGTGCCTCTTCCAGGCCCTGTCTCA	1500
OY	1712	AACCATCT	CCAAATTAGTCCAACCCAGAGGGCTGGCACTGCCCATCATCATCTATGCTCT	1771
Db	1501	AACCATCT	CCAAATTAGTCCAACCCAGAGGGCTGGCACTGCCCATCATCATCTATGCTCT	1560
OY	1772	TGCTGCCA	GTGCCAATTAAAGGGCTGATTGGCCAACTGG	1811
Db	1561	TGCTGCCA	GTGCCAATTAAAGGGCTGATTGGCCAACTGG	1600

Db 1 GCTGCTCCCTGGTTGCTGGGTGCAGAAATGCTGGGTTCTGGGTTTCTGGATTTCCGGGACG 60

QY 232 TTCAACAGTAGCCTTGTGCGGCGCTCTCCCTGGGTGAATCGTCCGCGCGCGATGCCCCGGGAC 291

Db 61 TTCAACAGTAGCCTTGTGCGGCGCTCTCCCTGGGTGAATCGTCCGCGCGCGATGCCCCGGGAC 120

QY 292 GGCCTAGGCTCCGGGGGTCCGGGGGCCCAAGGATTCGGGCTGAGATTGA CGGGGATC 351

Db 121 GACCTAGGCTCCGGGGGTCCGGGGGCCCAAGGATTCGGGCTGAGATTGA CGGGGATC 180

QY 352 CGGATGACCGCGCGCCCGCGGCTTCA CGGACGGGTCCAGACTGTGGAGAAAG 411

Db 181 CGGATGACCGCGCGCCCGCGGCTTCA CGGACGGGTCCAGACTGTGGAGAAAG 240

QY 412 TCGCGGGACGGGTCCTTAGATGCCATGCTTAGAGCCAAAGATGCTTAGCTTAATAG 471

Db 241 TCGCGGGACGGGTCCTTAGATGCCATGCTTAGAGCCAAAGATGCTTAGCTTAATAG 300

QY 472 TGTGACCTACATATGACTTCACTCAGTTTGTGATCCGTAATAATGACAAATTCGAA 531

Db 301 TGTGACCTACATATGACTTCACTCAGTTTGTGATCCGTAATAATGACAAATTCGAA 360

QY 532 GCTACTCACATGCTGTGAGAGATTAAATGAACAATGCTTTGAAGCTCTTTGAC 591

Db 361 GCTACTCACATGCTGTGAGAGATTAAATGAACAATGCTTTGAAGCTCTTTGAC 420

RESULT 7				
CR597629				
LOCUS	CR597629	1596 bp	mRNA	linear
DEFINITION	full-length CDNA clone CSDBJ0021Y105 of B cells (Ramos cell line)			
	Cot 25-normalized of Homo sapiens (human).			

DB 421 GAGGAGGCTTCGAGACGAGGCTTCGCGGAGAGACA CTTGCTGTACCAAGGACACAC 480

OY 652 AGGAGCATGAAGACCCCGTGGAGCTGGCCGTCAGTGGAAATGCAAGACCTTCGACCTTTCA 711

ACCESSION	CR597629
VERSION	CR597629.1 GI:50478436
KEYWORDS	HTC; cNSLT cDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Exonates	Method: Characterization: Nucleotide Sequence:

Db 481 AGGCGAGATGAAGACCCCCCTGGAGCTGGCCGTCAGTGGAGTGCAGACCCCTCGGCTTCA 540

Qy 712 GCACCGCTGCCGAGGTGGCTCAACGGGTCAAGGCAAGACGTCATVATGGATGAGACTCT 771

Db 541 GCACCGCTGCCGAGGTGGCTCAACGGGTCAAGGCAAGACGTCATVATGGATGAGACTCT 600

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Oy 772 GTTTGAGCAGCCAGCAGGACCCGGCCTAACCCACAGCATTCGATCCGCGCTTGGTGGA 831  
 Db 601 GTTTGAGCAGCCAGCAGGACCCGGCCTAACCCACAGGATTCGATCCGCGCTTGGTGGA 660  
 Oy 832 GAAGGCTATTCGATCCGCGCTTGGAAGGAGGAGGATTCGATTCGCGCTTGGTGGA 891

REMARK Contact : Feng Liang Email : fliang@fitech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue  
2 (bases 1 to 1596)  
REFERENCE Genoscope.

Accession	Gene	Length (bp)
U00096.1	gagC	892
U00096.1	gagA	892
U00096.1	gagB	892
U00096.1	gagD	892
U00096.1	gagE	892
U00096.1	gagF	892
U00096.1	gagG	892
U00096.1	gagH	892
U00096.1	gagI	892
U00096.1	gagJ	892
U00096.1	gagK	892
U00096.1	gagL	892
U00096.1	gagM	892
U00096.1	gagN	892
U00096.1	gagO	892
U00096.1	gagP	892
U00096.1	gagQ	892
U00096.1	gagR	892
U00096.1	gagS	892
U00096.1	gagT	892
U00096.1	gagU	892
U00096.1	gagV	892
U00096.1	gagW	892
U00096.1	gagX	892
U00096.1	gagY	892
U00096.1	gagZ	892

TITLE	Direct Submission
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:secref@genoscope.cns.fr">secref@genoscope.cns.fr</a> - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

Accession	Sequence	Length
Dd	721 GAGCTGTGAGACCAACCCCTTAAGGGGAGGACACCCCACTTACACCAAGAGAGAGAA	780
Qy	952 CAAAATACAGACCCATCAGCCACACCCCGTCTTACTGTGATGAGTGCCTTTGGCTCCCG	1011
Dd	781 CAAAATACAGACCCATCAGCCACACCCCGTCTTACTGTGATGAGTGCCTTTGGCTCCCG	840

**FEATURES**  
location/Qualifiers

Accession	Sequence	Position
Oy	ATCTGAAGGCGCCAGCTTTCGGGGGCCCCGCGATGCGAGAAAGGGGGATGCGCAAAAGCTCCG	1012 1071
Dh	ATCTGAAGGCGCCAGCTTTCGGGGGCCCCGCGATGCGAGAAAGGGGGATGCGCAAAAGCTCCG	841 900

```

SOURCE
1. .1376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL002Y105"
/trisue_type="B cells ( Ramos cell line) Not 25-normalized"

```

QY	1072	IGCTCTCTGTGGAGCGGCACACCTACCCCCAGGGGTAAGCATTCTGCCCCCCCCAGGGA	1131
Db	901	TGCTCTTGTGGAGCGGCACCTACCCCCAGGGTAAGCATTCTGCCCCCCCCAGGGA	960
QY	1132	GGCACCATTGGGACCATTCACCCAGCTGGTCCCTCCAAAGA CAGAGCGCGGGGACAGCGGC	1191

ORIGIN	/plasmid=1"pcmwSPORT_6"
Query Match	88.1%; Score 1596; DB 3; Length 1596;
Best Local Similarity	100.0%; Pred. No. 0;

Db	961	GGCACCACTGGAGCCCAATTCACCAAGCTGGTCCCTCCAGACAGAGCCGGGGCCACAGCCGC	1020
Qy	1192	AGACTCCACAGAGTTATCTATGGGTGGATTACACTTTCACGCCCCCTGMAAGCGGGAGCT	1251
Db	1021	AGACTCCACAGAGTTATCTATGGGTGGATTACACTTTCACAGCCCCCTCGAAGCGGGAGCT	1080

	Matches	1596;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	172	GCTGCTCCCTGGTTGCTGGGCAAGTCTGGGTTCTGGGTTCTGAGTTCCGGGCGG	231							

1252 TTCCCATTCCTCACCACTGAATGTCCAGACATGATCATCCAGCACCAATGCCCC 1311

Db 1081 TTCACATTCCTCCACCCACCTGATATGTCCCAAGACCTGTCTCATCCAGCCACCAAGTGCACC 1140

Qy 1312 CCACAATATGGGCTCCAGGATCTGAGGCTTCCAGTAAAGGGTGAACCTTCCGAGCC 1371

Db 1141 CCACAATATGGGCTCCAGGATCTGAGGCTTCCAGTAAAGGGTGAACCTTCCGAGCC 1200

Qy 1372 CTTGTGATCTCCAGGCTCGCTGATGATGATTTCAAGTCCATTAACCCAGAGAG 1431

Db 1201 CTTGTGATCTCCAGGCTCGCTGATGATGATTTCAAGTCCATTAACCCAGAGAG 1260

Qy 1432 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGTTCCTA 1491

Db 1261 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGTTCCTA 1320

Qy 1492 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGTTCCTA 1551

Db 1321 TGGGGCCACCCAGAAACCAAGCCCCCTTGAAGATGATCTTTTCATCAGGGTTCCTA 1380

Qy 1552 AACCCAGGATTTAAAGAAAGCCCTGTGGGGCAGACAGATAGAGGGTGGGAGTG 1611

Db 1381 AACCCAGGATTTAAAGAAAGCCCTGTGGGGCAGACAGATAGAGGGTGGGAGTG 1440

Qy 1612 CTTCCCTTTATCTGACATCTCTAGTGTGATTTCTGCTTTTCTCCGATTCGGATT 1671

Db 1441 CTTCCCTTTATCTGACATCTCTAGTGTGATTTCTGCTTTTCTCCGATTCGGATT 1500

Qy 1672 GGGGGCCACCTTAAGATGCTCTCTCCAGCCCTGTCTTAACCATTAATTAATGTC 1731

Db 1501 GGGGGCCACCTTAAGATGCTCTCTCCAGCCCTGTCTTAACCATTAATTAATGTC 1560

Qy 1732 CAACCCAGGGGCTGGACCTCCCATCATCATTT 1767

Db 1561 CAACCCAGGGGCTGGACCTCCCATCATCATTT 1596

RESULT 8  
CR606890  
LOCUS  
DEFINITION full-length cDNA clone CL0BB006ZC02 of Neuroblastoma of Homo sapiens (human).  
ACCESSION CR606890  
VERSION CR606890.1 GI:50487697  
KEYWORDS HTC; cNSLT cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Paraday Avenue  
2 (bases 1 to 1575)  
REFERENCE  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
FEATURES  
source location/Qualifiers  
1..1575  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CL0BB006ZC02"  
/issue\_type="Neuroblastoma"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 86.0%; Score 1557; DB 3; Length 1575;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1575; Conservative 0; Mismatches 0; Indels 8; Gaps 1;  
Qy 213 TTTTGGATTTGCGGGGCGGTTACAGATGAGCTGTGCGGCTCTCGGGTGAAGTCCGTC 272  
Db 1 TTTTGGATTTGCGGGGCGGTTACAGATGAGCTGTGCGGCTCTCGGGTGAAGTCCGTC 60  
Qy 273 GCGCGCGGTCCCGGAGCGGCTAGGCTGCGGGGGTCCGGAGCCCAAGCATTCGGG 332  
Db 61 GCGCGCGGTCCCGGAGCGGCTAGGCTGCGGGGGTCCGGAGCCCAAGCATTCGGG 120  
Qy 333 CTGCAATTTGATGCGGGATTCGCGATGACCGGCGCCCGGCTTACCGAGGCTC 392  
Db 121 CTGCAATTTGATGCGGGATTCGCGATGACCGGCGCCCGGCTTACCGAGGCTC 180  
Qy 393 AGACCTGGTGGGAAGAGGTGCGGGAGCGGGTCCCTGAGGATCCCGATCCGAGGCA 452  
Db 181 AGACCTGGTGGGAAGAGGTGCGGGAGCGGGTCCCTGAGGATCCCGATCCGAGGCA 240  
Qy 453 AGATGCTCAGCTTTTAAAGTGTGACCTACATGTGACTTCACTCAGTTTGTGATCCG 512  
Db 241 AGATGCTC-----AGGTGTGACCTACATGTGACTTCACTCAGTTTGTGATCCG 292  
Qy 513 TAAATGGAACAATTGGAAGCTACTTACAGTGTCTGTGAGAGATTAAATGAACAATG 572  
Db 293 TAAATGGAACAATTGGAAGCTACTTACAGTGTCTGTGAGAGATTAAATGAACAATG 352  
Qy 573 CTTGTAAGCTCTTTGAGAGAGGAGCTCTGGAACAGAGGCTCGGCGGCAAGACAC 632  
Db 353 CTTGTAAGCTCTTTGAGAGAGGAGCTCTGGAACAGAGGCTCGGCGGCAAGACAC 412  
Qy 633 TGTCTGACACAGGAGCAACAGCAGACATGAAAGACCCCGTGAAGTGGCCGTCACTGAG 692  
Db 413 TGTCTGACACAGGAGCAACAGCAGACATGAAAGACCCCGTGAAGTGGCCGTCACTGAG 472  
Qy 693 TGCAGACCTTGGGCTTTAGACACCGCTGCGAGGTGGCTACCGGGTCAAGGCCAGAGCT 752  
Db 473 TGCAGACCTTGGGCTTTAGACACCGCTGCGAGGTGGCTACCGGGTCAAGGCCAGAGCT 532  
Qy 753 CATATGTGATGAGACTGCTGTTTGGACCCGACAGAGGACCCCGGCTTACCCGAGAT 812  
Db 533 CATATGTGATGAGACTGCTGTTTGGACCCGACAGAGGACCCCGGCTTACCCGAGAT 592  
Qy 813 TCGATCCGCTCGGTGAGAAAGCTTAACAGAACAGAGCGTGGCAAGAGGATCGA 872  
Db 593 TCGATCCGCTCGGTGAGAAAGCTTAACAGAACAGAGCGTGGCAAGAGGATCGA 652  
Qy 873 AGGCTTGGGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCAAGACCCCAACC 932  
Db 653 AGGCTTGGGGCAAAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCAAGACCCCAACC 712  
Qy 933 TCACACCAAGGAAGAAAGAAATTAACAGACCATTAAGACACACCCCGTTACTGTGATG 992  
Db 713 TCACACCAAGGAAGAAAGAAATTAACAGACCATTAAGACACACCCCGTTACTGTGATG 772  
Qy 993 AGTGCCTGTTTGGCTCCGATCTGAAGCGCAGCTTGGGGGCCCGGCGATGCGAAG 1052  
Db 773 AGTGCCTGTTTGGCTCCGATCTGAAGCGCAGCTTGGGGGCCCGGCGATGCGAAG 832  
Qy 1053 GGGATGCGCAAAAGCTCGTGTCTCTTGTGAGCGCACCACTAACCCCGAGGGTACGC 1112  
Db 833 GGGATGCGCAAAAGCTCGTGTCTCTTGTGAGCGCACCACTAACCCCGAGGGTACGC 892  
Qy 1113 ACTGCGCCCGCCCAAGAGAGACACATGCGAGACCATTAACCCCGAGGGTACGC 1172  
Db 893 ACTGCGCCCGCCCAAGAGAGACACATGCGAGACCATTAACCCCGAGGGTACGC 952  
Qy 1173 CAGAGCCGGGCGCAGCGCAGACTCCAGAGTTATCATGAGGTGAGTTACACTTTCAC 1232  
Db 953 CAGAGCCGGGCGCAGCGCAGACTCCAGAGTTATCATGAGGTGAGTTACACTTTCAC 1012

QY 1233 GCCCCTGAAGCGGGGACCTTCCATTCCTCACCCACTGATGTCCCGACGACTGGTC 1292  
 Db 1013 GCCCCTGAAGCGGGGACCTTCCATTCCTCACCCACTGATGTCCCGACGACTGGTC 1072  
 QY 1293 ATCCAGCCACGAGTCCCGCCACAAATGGGCTCTAGAGCTTCCAGCTTCCAGCTCAG 1352  
 Db 1073 ATCCAGCCACGAGTCCCGCCACAAATGGGCTCTAGAGCTTCCAGCTTCCAGCTCAG 1132  
 QY 1353 GGGGACCTTCCGAGGAGGCTTGGTGAATTCAGAGGCTCGCTCAGTTAGCATTTCACTGC 1412  
 Db 1133 GGGGACCTTCCGAGGAGGCTTGGTGAATTCAGAGGCTCGCTCAGTTAGCATTTCACTGC 1192  
 QY 1413 CATTCACCCGACGAGAGTGGGCGCAACCCAGAAACCAAGCCCTTGGAAATGATACT 1472  
 Db 1193 CATTCACCCGACGAGAGTGGGCGCAACCCAGAAACCAAGCCCTTGGAAATGATACT 1252  
 QY 1473 CTTTCATCAGGGTGGCTTATGGGCGCACGCGGACAGGTATGGCCCTTTCAGGGGTAGA 1532  
 Db 1253 CTTTCATCAGGGTGGCTTATGGGCGCACGCGGACAGGTATGGCCCTTTCAGGGGTAGA 1312  
 QY 1533 GGACATTCATCAGCCAGGGGAAACCCAGGTATTAAGAGCCCTTGGGCGGACAGAC 1592  
 Db 1313 GGACATTCATCAGCCAGGGGAAACCCAGGTATTAAGAGCCCTTGGGCGGACAGAC 1372  
 QY 1593 ATACAGAGGGGTGGGCAAGTGCCTCCCTTATCTGACATCTCTAGTCAATCTTGGCTTT 1652  
 Db 1373 ATACAGAGGGGTGGGCAAGTGCCTCCCTTATCTGACATCTCTAGTCAATCTTGGCTTT 1432  
 QY 1653 TTCTCCGATTTGGGATTTGGGCGCACCTCTAAGATGCTCTCTCAGCCCTGTCTCA 1712  
 Db 1433 TTCTCCGATTTGGGATTTGGGCGCACCTCTAAGATGCTCTCTCAGCCCTGTCTCA 1492  
 QY 1713 CCATCTCCAAATTAAGTGCACACCCAGGGCGTGGCACTCCGACATCATCATCTTCTT 1772  
 Db 1493 CCATCTCCAAATTAAGTGCACACCCAGGGCGTGGCACTCCGACATCATCATCTTCTT 1552  
 QY 1773 GCTGCCAAGTGCATTAACCGGC 1795  
 Db 1553 GCTGCCAAGTGCATTAACCGGC 1575

RESULT 9  
 CR625812 1560 bp mRNA linear HTC 21-JUN-2004  
 LOCUS full-length cDNA clone CS0D1019Y108 of Placenta Cot 25-normalized  
 DEFINITION of Homo sapiens (human).  
 ACCESSION CR625812  
 VERSION CR625812.1 GI:50506619  
 KEYWORDS HTC; CNSLT cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1560)  
 AUTHORS La,W.B., Gruber,C., Jesse,J., and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue  
 REFERENCE 2 (bases 1 to 1560)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage : BP 101 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 FEATURES Location/Qualifiers

source 1..1560  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1019Y108"  
 /issue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 85.1%; Score 1541; DB 3; Length 1560;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 1560; Conservative 0; Mismatches 0; Indels 9; Gaps 1;  
 QY 191 GTGGAAGTCTGGGTTCTGGGTTTCTGATTTGGGCGGCGCTTCAACGATGCTGTGCC 250  
 Db 1 GTGGAAGTCTGGGTTCTGGGTTTCTGATTTGGGCGGCGCTTCAACGATGCTGTGCC 60  
 QY 251 GGGTCTCTGGGTTAGTCCGTCCGCGCGGCTGCCCGGACCGGCTTACGCTGCCGGGGT 310  
 Db 61 GGGTCTCTGGGTTAGTCCGTCCGCGCGGCTGCCCGGACCGGCTTACGCTGCCGGGGT 120  
 QY 311 CCGGGGCCCCAGGACATTCGGGGTCTGAGATTGACCGGGATCCCGGATGACCCGCGGCC 370  
 Db 121 CCGGGGCCCCAGGACATTCGGGGTCTGAGATTGACCGGGATCCCGGATGACCCGCGGCC 180  
 QY 371 CCGGCGCTTCAACGAGCGGCTCAGACCTGGTGGGAAAGAGGTGCGGGGACGGGTCCCTGA 430  
 Db 181 CCGGCGCTTCAACGAGCGGCTCAGACCTGGTGGGAAAGAGGTGCGGGGACGGGTCCCTGA 240  
 QY 431 GGATCCGATGCTTACGAGCGCAAGATGCTCAGCTTTAAGTGTGACTTACATGTGAC 490  
 Db 241 GGATCCGATGCTTACGAGCGCAAGATGCTCAGCTTTAAGTGTGACTTACATGTGAC 300  
 QY 491 TTGACCTCAGTTTGTGATCCGTAATAAGCAAAATTCGAAGTACTTACAGTGTGTT 550  
 Db 301 TTGACCTCAGTTTGTGATCCGTAATAAGCAAAATTCGAAGTACTTACAGTGTGTT 350  
 QY 551 GAGAGATTAAATGAACAATGCTTTGAAAGCTTTTGGAGAGGAGCCTCGGAAGCAG 610  
 Db 361 GAGAGATTAAATGAACAATGCTTTGAAAGCTTTTGGAGAGGAGCCTCGGAAGCAG 420  
 QY 611 GGCCTGGCGGCGAGACACCTGCTGTCAACGGGACCAAGCGCATGAAGACCCCG 670  
 Db 421 GGCCTGGCGGCGAGACACCTGCTGTCAACGGGACCAAGCGCATGAAGACCCCG 480  
 QY 671 GTGAGCTGGCGGTGATGAGGATGAGACCTCGCTTCAAGACCGCTGCGAGGTGGC 730  
 Db 481 GTGAGCTGGCGGTGATGAGGATGAGACCTCGCTTCAAGACCGCTGCGAGGTGGC 540  
 QY 731 TACCGGATCAAGGCGCAGACGTCATATGAGATGAGACTCTGTTGGAGCCGACAGGC 790  
 Db 541 TACCGGATCAAGGCGCAGACGTCATATGAGATGAGACTCTGTTGGAGCCGACAGGC 660  
 QY 791 ACCCGGCTTACCCGACCGGACTTGCATCCGCTTGGGTGAGAGGCTTACGACACGA 850  
 Db 601 ACCCGGCTTACCCGACCGGACTTGCATCCGCTTGGGTGAGAGGCTTACGACACGA 660  
 QY 851 GCGGTGGGCAAGGAGCATCGAAGGCTTGGGGGCAAGGGGAGCTGTAGCAACACCC 910  
 Db 661 GCGGTGGGCAAGGAGCATCGAAGGCTTGGGGGCAAGGGGAGCTGTAGCAACACCC 720  
 QY 911 TCAGGGGAGACACCCGACCCCTCACACCAAGAGAGAAACAAATACAGACCCATCAGC 970  
 Db 721 TCAGGGGAGACACCCGACCCCTCACACCAAGAGAGAAACAAATACAGACCCATCAGC 771  
 QY 971 CACACCCGCTTCTACTGTGATGATGCTGCTGTTGCTCCGATCTGAAGCGCGACCTTC 1030  
 Db 772 CACACCCGCTTCTACTGTGATGATGCTGCTGTTGCTCCGATCTGAAGCGCGCGCTTC 831  
 QY 1031 GGGGCGCGCGGATGCGGAAGGGGAGTGCAGGAAGCTCGGTCTCTTGTGAGACCA 1090  
 Db 832 GGGGCGCGCGGATGCGGAAGGGGAGTGCAGGAAGCTCGGTCTCTTGTGAGACCA 891

QY	1091	CCACCTTACCCCCAGGGGGTAGGCACTGCGCCCCGGCCCCAGAGAGAGCAACACTGCGAGCAATT	1159
Db	892	CCACCTACCCCCAGGGGGTAGGCACTGCGCCCCGGCCCCAGAGAGAGCAACACTGCGAGCAATT	951
QY	1151	CACCCAGCTGGTCCCTCCAGAAGAGAGCCGGGGCCAGCGGAGACTCCAGAAATTATCT	1210
Db	952	CACCCAGCTGGTCCCTCCAGAAGAGAGCCGGGGCCAGCGGAGACTCCAGAAATTATCT	1011
QY	1211	ATGGGTGGGGTTACACTCTTCAAGCCCCCTGAAAGGGGGGACCTTCCCATTTCCCTCAACCAC	1270
Db	1012	ATGGGTGGGGTTACACTCTTCAAGCCCCCTGAAAGGGGGGACCTTCCCATTTCCCTCAACCAC	1071
QY	1271	CTGAATGTCCTCCAGCAGCTGTATCTCAAGCCACAAGTGGCCCCCACAACAATAGGGCTCAG	1330
Db	1072	CTGAATGTCCTCCAGCAGCTGTATCTCAAGCCACAAGTGGCCCCCACAACAATAGGGCTCAG	1131
QY	1331	GATCTCAGAGCCCTTCCAGCTCAGGGGTGACCTTCCGAGACCCCCCTGTGTACTTCCAGGGCT	1390
Db	1132	GATCTCAGAGCCCTTCCAGCTCAGGGGTGACCTTCCGAGACCCCCCTGTGTACTTCCAGGGCT	1191
QY	1391	CGCTCAGTTAGCATTTTCAGTGCCATCTACCCCAAGACAGAGGTGGGGCCACCCAGAAACA	1450
Db	1192	CGCTCAGTTAGCATTTTCAGTGCCATCTACCCCAAGACAGAGGTGGGGCCACCCAGAAACA	1251
QY	1451	AAGCCCCCTTGGAAATGATTACTCTTTCATCAGGGTTGCTTATGSGGGCCACGGGCAGAGT	1510
Db	1252	AAGCCCCCTTGGAAATGATTACTCTTTCATCAGGGTTGCTTATGSGGGCCACGGGCAGAGT	1311
QY	1511	ATGGCCCCCTTGCAGGGGTAGAGGACATTCAACCCAGAGGGAACCCAGATTTAAAGAA	1570
Db	1312	ATGGCCCCCTTGCAGGGGTAGAGGACATTCAACCCAGAGGGAACCCAGATTTAAAGAA	1371
QY	1571	GCCCCCTGTGGGGGAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTATATCTGTCAAA	1630
Db	1372	GCCCCCTGTGGGGGAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTATATCTGTCAAA	1431
QY	1631	TCTCTAGTCGATTTCTGCTCTTTTCTCCCAATTCGGATTTGGGGGGCCACTCTAAATAG	1690
Db	1432	TCTCTAGTCGATTTCTGCTCTTTTCTCCCAATTCGGATTTGGGGGGCCACTCTAAATAG	1491
QY	1691	CCTCTCTTCAGACCCTGTCTCAACCATATCTCAAAATTAGTGCCAACCCAGGGGCTGTGCAC	1750
Db	1492	CCTCTCTTCAGACCCTGTCTCAACCATATCTCAAAATTAGTGCCAACCCAGGGGCTGTGCAC	1551
QY	1751	CTCCCACT 1759	
Db	1552	CTCCCACT 1560	

RESULT 10			
CR617377		1488 bp	mRNA linear HTC 21-JUL-2004
LOCUS			
DEFINITION	full-length cDNA clone CSDDC024YF06 of Neuroblastoma Cot 25-normalized of Homo sapiens (human).		
ACCESSION	CR617377		
VERSION	CR617377.1	GI:50498184	
KEYWORDS	HTC; CNSLT cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Li,W.B., Gruber,C., Jesssee,J. and Polayes,D. Full-length cDNA libraries and normalization unpublished		
REFERENCE	Contact : Feng Liang Email : fliang@lifeitech.com URL :		
AUTHORS	<a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> Invitrogen Corporation 1600		
TITLE	Paradise Avenue		
JOURNAL	2 (bases 1 to 1488) Genoscope. Direct Submission Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:seqref@genoscope.cns.fr">seqref@genoscope.cns.fr</a> )		
REMARK			

COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen
FEATURES	Location/Qualifiers
source	1..1488
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="CS0DC024YF06"
	/tissue_type="Neuroblastoma Cot 25-normalized"
	/plasmid="pCMVSPORT_6"
ORIGIN	
Query Match	74.3%; Score 1346; DB 3; Length 1488;
Best Local Similarity	91.9%; Pred. No. 0;
Matches 1488; Conservative	0; Mismatches 0; Indels 132; Gaps 1;
177	TCCTGTGCTGGGTGCAAGTCTGGGTTCTGGGTTCTGATTGCGGGCCGTTAC
1	TCCCTGGTGTCTGGGTGCAAGTCTGGGTTCTGGGTTCTGATTGCGGGCCGTTAC
237	ACGTAGCTGTGCTGGGCTCTCTGGGTGAGTCCGTCCGCGCGCGGTGCCGGAGCGCT
61	ACGTAGCTGTGCTGGGCTCTCTGGGTGAGTCCGTCCGCGCGCGGTGCCGGAGCGCT
297	AGGCTGCGGGGGGTCCGGGGGCCAGGAGATTCCGGGGCTGCAGATTGACGGGGATCCGGA
121	AGGCTGCGGGGGGTCCGGGGGCCAGGAGATTCCGGGGCTGCAGATTGACGGGGATCCGGA
357	TGCACCGGCGCCCCCGCGCCCTTCACGACGAGGCTCCAGACCTGGTGGGAAGAGTCCG
181	TGCACCGGCGCCCCCGCGCCCTTCACGACGAGGCTCCAGACCTGGTGGGAAGAGTCCG
417	GGAAGGGTCCCTGAGATCCCGATGCTTACGAGCCAAAGATGCTTATAGTGTGA
241	GGAAGGGTCCCTGAGATCCCGATGCTTACGAGCCAAAGATGCTTATAGTGTGA
477	CCTACATGATGATTCACCTCACTTTTGTGATCCGTAATGACAAATTCCAGAGTAC
285	-----
537	TTCAAGTGTCTGTGAGAGATTAAATGAAACAATGCTTGTAAAGCTTTGCAGAGAG
285	-----AGGG
597	AGCTTCGAGACAGAGGCTTGCGCGGACAGACACTGCTGTCAACAGGACCAAGGCA
289	AGCTTCGAGACAGAGGCTTGCGCGGACAGACACTGCTGTCAACAGGACCAAGGCA
657	GCAAGAAAGACCCCGGTGAGACTGGCCGCTCATGTGGGAATGACACCTCGGCTTCAAGAC
349	GCAAGAAAGACCCCGGTGAGACTGGCCGCTCATGTGGGAATGACACCTCGGCTTCAAGAC
717	GCTCCGAGGTGTCTACCGGCTCAAGGACAGAGCTATATGTGATGAGACTCTGTTTG
409	GCTCCGAGGTGTCTACCGGCTCAAGGACAGAGCTATATGTGATGAGACTCTGTTTG
777	GCAAGCCAGACAGGCACTCGGCTTACCCACCGGACTTGCATCCGCTCTGGGTGGAGAG
469	GCAAGCCAGACAGGCACTCGGCTTACCCACCGGACTTGCATCCGCTCTGGGTGGAGAG
837	CTAACAGAAACAGAGGCTGTGGGCAAGAGGATTCGAAGGCTTTGGGGGCAAAAGGGAGCT
529	CTAACAGAAACAGAGGCTGTGGGCAAGAGGATTCGAAGGCTTTGGGGGCAAAAGGGAGCT
897	GTGAGACCAACCCCTCAAGGGGAGAGACCCCAACCTTCACACCAAGGAGAGAAACAAAT
589	GTGAGACCAACCCCTCAAGGGGAGAGACCCCAACCTTCACACCAAGGAGAGAAACAAAT
957	ACAACCCATCAGCAACCCCGTCTTACTGTGATGATGTCGCTTTGCTCCGATCTG



Db	649	ACAGACCATTAGGCGACACCCCGTTTACTGTGATGAGTCGCTGTTTGACTCCGATCTTG	1076
OY	1017	AAGCGCGAGCTTCGGGGCCCCCGCGGATGCGAAGGGAGATGCGCAAAAGCTCCGTGCTC	1076
Db	709	AAGGCGCAGACTTCGGGGCCCCCGCGGATGCGAAGGGAGATGCGCAAAAGCTCCGTGCTC	768
OY	1077	TCTTGTGACGCGACACACTACTACCCCGAGGGGTAGCACTGCGCCCGCCCGGAGGAGCGAC	1136
Db	769	TCTTGTGACGCGACACACTACTACCCCGAGGGGTAGCACTGCGCCCGCCCGGAGGAGCGAC	828
OY	1137	CACCTGCGAGCAATTCACCCAGCTGCTCCCTCCACAGACAGAGCCGCGGCGACAGACT	1196
Db	829	CACCTGCGAGCAATTCACCCAGCTGCTCCCTCCACAGACAGAGCCGCGGCGACAGACT	888
OY	1197	CCGAGAAATTATCTATGCGGTGGGTAACTCTTACCGCCCCCTGAAGGGGGAATTTTCCC	1256
Db	889	CCGAGAAATTATCTATGCGGTGGGTAACTCTTACCGCCCCCTGAAGGGGGAATTTTCCC	948
OY	1257	ATTCCCTCACCCACCTGATATGCCCCACACACTGATCTCCAGCAGCAGCAGTCCGCCACA	1316
Db	949	ATTCCCTCACCCACCTGATATGCCCCACACACTGATCTCCAGCAGCAGCAGTCCGCCACA	1008
OY	1317	CAATGCGCCTTAGATCTCAGGCTTTCACAGTCAAGGGTAACTTCCGAGCCCCCTGG	1376
Db	1009	CAATGCGCCTTAGATCTCAGGCTTTCACAGTCAAGGGTAACTTCCGAGCCCCCTGG	1068
OY	1377	TGACTCTCAGGGCTGCTCAGTTAGCAATTTACAGTCCATCTACCCCAAGAGAGGTGGG	1436
Db	1069	TGACTCTCAGGGCTGCTCAGTTAGCAATTTACAGTCCATCTACCCCAAGAGAGGTGGG	1128
OY	1437	CCACCCAGAAAACCAAAAGCCCCCTTGGAAATGATACTCTTTTCATCAGGGTTTCCATG	1496
Db	1129	CCACCCAGAAAACCAAAAGCCCCCTTGGAAATGATACTCTTTTCATCAGGGTTTCCATG	1188
OY	1497	CCAGGCGCAGAGGTATGCGCCCTTGGCCAGGGTAGAGAGACATTCATCACCCAGAGAAACC	1556
Db	1189	CCAGGCGCAGAGGTATGCGCCCTTGGCCAGGGTAGAGAGACATTCATCACCCAGAGAAACC	1248
OY	1557	CAGGTATTAAGAAGCCCCCTGTGGGGGAGACAGACATAGCAGGGGTGGCAGTGCTCC	1616
Db	1249	CAGGTATTAAGAAGCCCCCTGTGGGGGAGACAGACATAGCAGGGGTGGCAGTGCTCC	1308
OY	1617	CTTTATCTGACAACTCTTAGTCGATTTCTTGCCCTTTTCTCCGATTCGGATTTGGGGG	1676
Db	1309	CTTTATCTGACAACTCTTAGTCGATTTCTTGCCCTTTTCTCCGATTTGGGGG	1368
OY	1677	CCACCTCAAAATGCGCTCTCTCCAGCCCTGTCTCAACATCTCAATTAATGATGCAAC	1736
Db	1369	CCACCTCAAAATGCGCTCTCTCCAGCCCTGTCTCAACATCTCAATTAATGATGCAAC	1428
OY	1737	CAGGGGCTGGGACCTCCCAATCATATCCATTTGCTCTGTCGCAAGTGGAAATTAACGGCG	1796
Db	1429	CAGGGGCTGGGACCTCCCAATCATATCCATTTGCTCTGTCGCAAGTGGAAATTAACGGCG	1488
RESULT 11			
LOCUS	CR622791	1467 bp	mRNA
DEFINITION	full-length cDNA clone CSD0L050Y24 of B cells (Ramos cell line)		
ACCESSION	CR622791		
VERSION	CR622791.1	GI:50503598	
KEYWORDS	HTC; CNSLT_cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Li, W. B., Gruber, C., Jessee, J. and Polayres, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished		
REMARK	Contact : Peng Liang Email : filiang@lifetech.com URL :		

REFERENCE	AUTHORS	TITLE	JOURNAL
http://fulllength.invitrogen.com/InvitrogenCorporation1600	Parady Avenue	2 (bases 1 to 1467)	Genoscope.
Genoscope.	Direct Submission	Submitted (20-JUL-2004)	Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr			
- Web : www.genoscope.cns.fr			
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers			
end enriched, double-strand cDNA was digested with Not I and cloned			
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library			
was normalized. Library was constructed by Life technologies, a			
division of Invitrogen.			
location/qualifiers			
1. 1467			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="CS0DL005Y024"			
/tissue_type="B cells (Ramos cell line) Cot 25-normalized			
/plasmid="pCMVSPORT_6"			
ORIGIN			
Query Match	73.2%	Score 1325;	DB 3; Length 1467;
Best Local Similarity	91.7%	Pred. No. 0;	
Matches 1467; Conservative	0;	Mismatches	0; Indels 132; Gaps 1;
177	TCCCTGGTGTGCGGATGCAAGATCGGGTCTGAGTTCTGAGATTGCGGGCGGCTTAC	236	
1	TCCCTGGTGTGCGGATGCAAGATCGGGTCTGAGTTCTGAGATTGCGGGCGGCTTAC	60	
237	ACGTAGCTTAGCGGCTCTCGGATGATCCGTCCGCGCGCGGTCGCCGAGCGCT	296	
61	ACGTAGCTTAGCGGCTCTCGGATGATCCGTCCGCGCGCGGTCGCCGAGCGCT	120	
297	AGGCTGCGCGGGGTCGGGGGCCCAAGGATTCGGGCTGCGAGATTGACGGGATTCGGGA	356	
121	AGGCTGCGCGGGGTCGGGGGCCCAAGGATTCGGGCTGCGAGATTGACGGGATTCGGGA	180	
357	TGCACCGCGCCCGCCCGCGCTTACAGCAGACGGATTCAGACTGGTGGAGAAAGTGGG	416	
181	TGCACCGCGCCCGCCCGCGCTTACAGCAGACGGATTCAGACTGGTGGAGAAAGTGGG	240	
417	GGACGGGTCCCTGAGATCCGATGCTTACGAGCGAAGATCTGACTTATAGGTGTA	476	
241	GGACGGGTCCCTGAGATCCGATGCTTACGAGCGAAGATCTGACTTATAGGTGTA	284	
477	CCTACATGATGATTCACCTCAAGTTTGTGATCCGTAATGACAAATTGCAAGCTAC	536	
285	-----	284	
537	TTACAGATGCTGTTAGAGATTAATGAAACAATGCTTTGTAAGCTTTTGAGAGAGG	596	
285	-----AGGG	288	
597	AGCTTCGAGACAGAGGCTTGGCGCGGAGAGACACTGCTGTCAACGAGACCAAGGCA	656	
289	AGCTTCGAGACAGAGGCTTGGCGCGGAGAGACACTGCTGTCAACGAGACCAAGGCA	348	
657	GCATGAAGACCCCGTGGAGTGGCGGTCAAGTGGGATGACAGACCTCGGCTTACGACCC	716	
349	GCATGAAGACCCCGTGGAGTGGCGGTCAAGTGGGATGACAGACCTCGGCTTACGACCC	408	
717	GCTGCGGAGGTGGTCAACGGGTCAAGGCAAGACGTCAATATGTGATGAGACTTGTGTTG	776	
409	GCTGCGGAGGTGGTCAACGGGTCAAGGCAAGACGTCAATATGTGATGAGACTTGTGTTG	468	
777	GCAGCCAGACAGGACCCCGGCTTACCCCAACGAGCTTGCATCCGCTTGGGTGAGAAAG	836	
469	GCAGCCAGACAGGACCCCGGCTTACCCCAACGAGCTTGCATCCGCTTGGGTGAGAAAG	528	
837	CTAAACAGACCAAGGCGTGGGCAAGAGGATGAAGGCTTTGGGGCAAAAGGAGACT	896	

Db	529	TTAA CAGAA CCA GAG CGT GGG CAA GAG GCA TGA AG CCG TTG GGG CCA AAG GGA GCT	588
Qy	897	GTG AGA CCA CCCC CT CAA GGG GCA GCA CCCC CA CCT CTA CAC CAA GGA AAG AA CAA T	956
Db	589	GTG AGA CCA CCCC CT CAA GGG GCA GCA CCCC CA CCT CTA CAC CAA GGA AAG AA CAA T	648
Qy	957	ACAGA CCA TGA GGC A GAC CCG GTC TTA CTG TGA TGA TGC GCT GTT TGG CTT CCG A TCT G	1016
Db	649	ACA GAC CCA TGA GGC A GAC CCG GTC TTA CTG TGA TGA TGC GCT GTT TGG CTT CCG A TCT G	708
Qy	1017	AAG GCG CAG CTT CG GGG CCCC CG GAG ATG CGA AAG GGG AATG CG CAA AG C TCG TGT C	1076
Db	709	AAG GCG CAG CTT CG GGG CCCC CG GAG ATG CGA AAG GGG AATG CG CAA AG C TCG TGT C	768
Qy	1077	TCT TGTG ACG CCA CCA CCA CTTA CCCC CAG GGG TGA GCA CTA TGC CCCC CG CCA GGA AG CAC	1136
Db	769	TCT TGTG ACG CCA CCA CCA CTTA CCCC CAG GGG TGA GCA CTA TGC CCCC CG CCA GGA AG CAC	828
Qy	1137	CAC TGG CCA CCA TTA CCA CCA CTA GCT GGT GCT CTA A GAC A GAC CCG GGG CCA GCG CACA CT	1196
Db	829	CAC TGG CCA CCA TTA CCA CCA CTA GCT GGT GCT CTA A GAC A GAC CCG GGG CCA GCG CACA CT	888
Qy	1197	CCAGA AGT TATCTA TGG TGG GTT A CACT CTTT CAC GCG CCCC CTG AAG GGG GGA CTTT TCC	1256
Db	889	CCAGA AGT TATCTA TGG TGG GTT A CACT CTTT CAC GCG CCCC CTG AAG GGG GGA CTTT TCC	948
Qy	1257	ATT CCT CTA CCA CCA CCA TGA TGT CCCC CAG CACT GGT CAT CAG CCA CCA GTC CCCC CACA	1316
Db	949	ATT CCT CTA CCA CCA CCA TGA TGT CCCC CAG CACT GGT CAT CAG CCA CCA GTC CCCC CACA	1008
Qy	1317	CAAT TGG GGC CTA CAG ATCT CAG GGC CTTT CCA GT CAG GGG TGA CTTT CCG A GCCC CTT GGG	1376
Db	1009	CAAT TGG GGC CTA CAG ATCT CAG GGC CTTT CCA GT CAG GGG TGA CTTT CCG A GCCC CTT GGG	1068
Qy	1377	TGACT CTA CAG GGC CTG CTA GTT GAC ATTT CAG TGC CAT CTA ACC CAA GAC GAG GTG GGG	1436
Db	1069	TGACT CTA CAG GGC CTG CTA GTT GAC ATTT CAG TGC CAT CTA ACC CAA GAC GAG GTG GGG	1128
Qy	1437	CCACC CAG AAA CCAA AGCCCC CTTT GAA AAT GAT A CTTT CAT CAG GGT TGC CTTAT GGGG	1496
Db	1129	CCACC CAG AAA CCAA AGCCCC CTTT GAA AAT GAT A CTTT CAT CAG GGT TGC CTTAT GGGG	1188
Qy	1497	CCAG GCG A CAG GTAT GGC CCCC TTG CCA GGG TGA GGA CATT CAT CACC CAG GAA ACC	1556
Db	1189	CCAG GCG A CAG GTAT GGC CCCC TTG CCA GGG TGA GGA CATT CAT CACC CAG GAA ACC	1248
Qy	1557	CAG GAT TTA AAG AAG CCCC CTG TGG GGG CAG A CAG CAT TGA GGA GGG GTG GCG AGT CCG	1616
Db	1249	CAG GAT TTA AAG AAG CCCC CTG TGG GGG CAG A CAG CAT TGA GGA GGG GTG GCG AGT CCG	1308
Qy	1617	CTTAT TCC TGA CAA TCT CTA TGC ATTT CTT C CTTT TTT C CCA TTT GCG A TTT GGG GGG	1676
Db	1309	CTTAT TCC TGA CAA TCT CTA TGC ATTT CTT C CTTT TTT C CCA TTT GCG A TTT GGG GGG	1368
Qy	1677	CCACC TTA GAT GCT CTT CTA GGC C CTT GCT CTA CCA CCA TTA CTA CCA ATTAT GTC CAA CC	1736
Db	1369	CCACC TTA GAT GCT CTT CTA GGC C CTT GCT CTA CCA CCA TTA CTA CCA ATTAT GTC CAA CC	1428
Qy	1737	CAG GGG C CTA CCA CCA CCA TTA CCA TTA CCA TTA CCA TTA CCA TTA CCA TTA CCA TTA CCA	1775
Db	1429	CAG GGG C CTA CCA CCA CCA TTA CCA TTA CCA TTA CCA TTA CCA TTA CCA TTA CCA TTA CCA	1467

RESULT	12
CR605763	
LOCUS	CR605763 1465 bp mRNA linear HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CS0PC020YCO8 of Neuroblastoma Cot 25-normalized of Homo sapiens (human).
ACCESSION	CR605763
VERSION	CR605763.1 GI:50486570
KEYWORDS	HTC; CNSLT_cDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK
REFERENCE	1	(bases 1 to 1465)		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homimidae; Homo.
REFERENCE	2	(bases 1 to 1465)		Li, W.B., Gruber, C., Jeesee, J., and Polayes, D. Full-length cDNA libraries and normalization unpublished
COMMENT				Contact : Feng Liang Email : fliang@life tech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue (bases 1 to 1465) Genoscope.
FEATURES				Direct Submission Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five primers end enriched, double-strand cDNA was dished with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
ORIGIN				Location/Qualifiers 1..1465 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODC020YC08" /issue_type="Neuroblastoma Cot 25-normalized" /plasmid="pCMVSPORT_6"
Query Match	73.1%	Score 1323;	DB 3;	Length 1465;
Best Local Similarity	91.7%;	Pred. No. 0;		
Matches 1465;	Conservative 0;	Mismatches 0;	Indels 132;	Gaps 1
QY	179	CCGCGTGGCTGGGGGCAAGGCTGGGGCTGGGGTTCTGGATTGCGGGGCGGTCACAC	238	
DB	1	CCGTGGTGGCTGGGGGCAAGGCTGGGGCTGGGGTTCTGGATTGCGGGGCGGTCACAC	60	
QY	239	GTACGCTGTGCGCGGCTCCTCGGCTGAGTCGCTCGCGCGGAGTCCCGGAGCGGCTAG	298	
DB	61	GTACGCTGTGCGCGGCTCCTCGGCTGAGTCGCTCGCGCGGAGTCCCGGAGCGGCTAG	120	
QY	299	GCTGCGCGGGGCTCGGGGCGCCGAGGCATTCCGGGCTGAGATTGACGCGGAGTCCCGAGT	358	
DB	121	GCTGCGCGGGGCTCGGGGCGCCGAGGCATTCCGGGCTGAGATTGACGCGGAGTCCCGAGT	180	
QY	359	CACCGCGCGCGCCCGCGCGGCTTACCGAGCGGCTCCAGACTGTGGGAGAGAGTCCGGG	418	
DB	181	CACCGCGCGCGCCCGCGCGGCTTACCGAGCGGCTCCAGACTGTGGGAGAGAGTCCGGG	240	
QY	419	ACGGGCTCCCTGAGGATCCCGATGCTGCGAGCAAGATGCTCAGCTTATAGGTGAC	478	
DB	241	ACGGGCTCCCTGAGGATCCCGATGCTCAGAGCAAGATGCTC-----	282	
QY	479	TACACATGTGACTTCACTCAGTTTGTGATCGTAAATGAGCAAAATTCGAGCTACTT	538	
DB	283	-----	282	
QY	539	CACAGTCTGTGAGAGGATTTAAATGAACAATGCTTTGTAAGCGGAG	598	
DB	283	-----AGGAG	288	
QY	599	CCTCGAAGACGGGCTGGCGCGGAGAGCAACCTGCTGTGACACGAGGACGACGAGC	658	
DB	289	CCTCGAAGACGGGCTGGCGCGGAGAGCAACCTGCTGTGACACGAGGACGACGAGC	348	
QY	659	ATGAAGAAGACCCCGTGAAGCTGGCGCTGAGTGGAGTGCAGACCTTCGAGCAGCAGC	718	
DB	349	ATGAAGAAGACCCCGTGAAGCTGGCGCTGAGTGGAGTGCAGACCTTCGAGCAGCAGC	408	
QY	719	TGCGGAGGTGGCTACCGGCTCAAGGCGGACGTCATATGTGATGAGACTCTGTTGGC	778	
DB	409	TGCGGAGGTGGCTACCGGCTCAAGGCGGACGTCATATGTGATGAGACTCTGTTGGC	468	





OY	1242	AGCGGGGACTTTCCATTCCCTCAACCCACTGAAATGTCGCCAGCATCTGATATCCAGCA	130
Db	481	AGCGGGGACTTTCCATTCCCTCAACCCACTGAAATGTCGCCAGCATCTGATATCCAGCA	540
OY	1302	CCAGTGGCCCCCAACAATATGGGCTCAGATCTCAGGCTTCACTCAAGGAGTGACT	136
Db	541	CCAGTGGCCCCCAACAATATGGGCTCAGATCTCAGGCTTCACTCAAGGAGTGACT	600
OY	1382	TCCGGAGCCCCCTGGTGACTTTCAGAGGCTCGCTCAAGTTAGCATTTTCAGTGCATCTAC	1421
Db	601	TCCGGAGCCCCCTGGTGACTTTCAGAGGCTCGCTCAGTTAGCATTTTCAGTGCATCTAC	660
OY	1422	CACGACGAGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATATCTTTTATCA	1481
Db	661	CACGACGAGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATATCTTTTATCA	720
OY	1482	GGGTTGCCTATGGGGCCACAGCGGACAGGTATGGCCCTTGGCCAGGGTATGAGAGCATTC	1541
Db	721	GGGTTGCCTATGGGGCCACAGCGGACAGGTATGGCCCTTGGCCAGGGTATGAGAGCATTC	780
OY	1542	TCACCCAGGGAACCCCAAGGTATTTAAAGAAAGCCCTGTGGGGGACAGACATAGCAGGG	1601
Db	781	TCACCCAGGGAACCCCAAGGTATTTAAAGAAAGCCCTGTGGGGGACAGACATAGCAGGG	840
OY	1602	GTGGGCAATGCTCCCTTATTCCTGAACAATCTATAGTGAATTTCTTGCTTTTCTCCCGA	1661
Db	841	GTGGGCAATGCTCCCTTATTCCTGAACAATCTATAGTGAATTTCTTGCTTTTCTCCCGA	900
OY	1662	TTGCGGATTTGGGGGCCACTCTGAAGATGSCCTCTCCAGCCCTGTCAACCAATATCC	1721
Db	901	TTGCGGATTTGGGGGCCACTCTGAAGATGSCCTCTCCAGCCCTGTCAACCAATATCC	960
OY	1722	AAATTATGTGCCAACCCAGGGGCTTGGCACTTCCCAATATTCATTTGTCTTGCTGCCAAG	1781
Db	961	AAATTATGTGCCAACCCAGGGGCTTGGCACTTCCCAATATTCATTTGTCTTGCTGCCAAG	1020
OY	1782	TGCGAATTAAGGCGTGATTGCCAA	1806
Db	1021	TGCGAATTAAGGCGTGATTGCCAA	1045

RESULT 14	
BX358473	
LOCUS	BX358473 1098 bp mRNA linear EST 08-APR-2004
DEFINITION	BX358473 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS001039YB15 5-PRIME, mRNA sequence.

ORGANISM	Homo sapiens
SOURCE	Homo sapiens (human)
KEYWORDS	EST.
VERSION	358473.2
ACCESSION	GI:4628390
BA358473	

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1098)
TITLE	Li, W. B., Gruber, C., Jessee, J., and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished (2001)
	On May 2, 2003 this sequence version replaced gi:30347985.

Contact: Genoscope  
Genoscope - Centre National de Séquençage  
2 rue Gaston Crémieux, CP 5706 - 91057 Evry cedex - FRANCE  
Email: [seger@genoscope.cns.fr](mailto:seger@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-ligo (dT) primer. Five primers  
end enriched, double-strand cDNA was digested with NotI and cloned  
into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized, library was constructed by Life technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
6578

For more information about this cluster, see  
<http://www.genoscope.cns.fr/cda?c=CE0D1039CA08QD1&c=6579.r>.

```
FEATURES
source
location/Qualifiers
1. 1098
/organism="Homo sapiens"
```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1039YB15"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-colligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

Query Match	55.9%;	Score 1012.2;	DB 5;	Length 1098;
Best Local Similarity	98.4%;	Pred. No. 1.7e-247;		
Matches 1060; Conservative	4;	Mismatches 8;	Indels 5;	Gaps 4

QY	196	AAGTCTGGGTTCTGGGTTCTCGAATTGCGCGGCGGTTCAACGTAAGCCTGTGCCGCTC	255
Db	18	AAGTCTGGGTTCTGGGTTCTGGAATTGCGCGGCGGTTCAACGTAAGCCTGTGCCGCTC	77
QY	256	CTCGGGTAATCCGTTCGCGCGCGGTCGCCCGGACCGGCTTAAGCTGCCGGGGTCCGGG	315
Db	78	CTCGGGTAAGTCCGTTCGCGCGCGGTCGCCCGGACCGGCTTAAGCTGCCGGGGTCCGGG	137
QY	316	GCCCCAGGCATTCGGGGCTGCGAGATTGACGCGGGATCCGGAATGACCCGCGGCCCCCGG	375
Db	138	GCCCCAGGCATTCGGGGCTGCGAGATTGAC - GGGAATCCGGAATGACCGGCGCCCCCGG	196
QY	376	CCCTCACCGACGGGTCCAGACCTGTGCGAAGAGGTGCGGGACGGGTCCCTGAGATC	435
Db	197	CCCTCACCGACGGGTCCAGACCTGTGCGAAGAGGTGCGGGACGGGTCCCTGAGATC	256
QY	436	CCGATGCTTAAGAGCGCAAGATGCTCAGTTTATAGGTGACTTAACAATGTGACTTTCAC	495
Db	257	CCGATGCTTAAGAGCGCAAGATGCTCAGTTTATAGGTGACTTAACAATGTGACTTTCAC	316
QY	496	CTCAGTTTGTGATCCGTAATAATGGACAATTCGAGCTACTTCACAGTCTGTTGAGAG	555
Db	317	CTCAGTTTGTGATCCGTAATAATGGACAATTCGAGCTACTTCACAGTCTGTTGAGAG	376
QY	556	GATTAAATGAACAATGCTTGTAAAGCTCTTTCGAGAGGAGCCCTCGGAAGCAGGCGCT	615
Db	377	GATTAAATGAACAATGCTTGTAAAGCTCTTTCGAGAGGAGCCCTCGGAAGCAGGCGCT	436
QY	616	GCGCGGCGAGACACACCTGCTGTCAACGAGGACCAACAGGACATGAAGAATCCCTGTGA	675
Db	437	GCGCGGCGAGACACACCTGCTGTCAACGAGGACCAACAGGACATGAAGAATCCCTGTGA	496
QY	676	GCTGGCCGTCAGTGGGATGACAGACCTCGGCTTCAGACCGCTGCGAGGTGCTACCG	735
Db	497	GCTGGCCGTCAGTGGGATGACAGACCTCGGCTTCAGACCGCTGCGAGGTGCTACCG	556
QY	736	GCTCAGGCGCAGACGTCATATGTGATGAGACTCTGTTGGCAGGCCGACGACGCCG	795
Db	557	GCTCAGGCGCAGACGTCATATGTGATGAGACTCTGTTGGCAGGCCGACGACGCCG	616
QY	796	GCCTAACCCCAACCGACCTTCGATCCGCCCTGGGTGAGAAAGGCTAACAAACAGAGGGGT	855
Db	617	GCCTAACCCCAACCGACCTTCGATCCGCCCTGGGTGAGAAAGGCTAACAAACAGAGGGGT	676
QY	856	GGGCAAGAGGACATCGAAGGCTTGGGGGCAAAAGGGAGCTGTGAGACACCCCTCTAAG	915
Db	677	GGGCAAGAGGACATCGAAGGCTTGGGGGCAAAAGGGAGCTGTGAGACACCCCTCTAAG	736
QY	916	GGGCAAGACCCCCACCTTCACACCAAGAGAAAGAACAAATACAGACCCATCAGCACAAC	975
Db	737	GGGCAAGACCCCCACCTTCACACCAAGAGAAAGAACAAATACAGACCCATCAGCACAAC	796
QY	976	CCCGCTTTACTGTATGATGCTGCTGTTGGCTCCCGAATCTAAGCGCGCAGCTTCGGGGC	1033
Db	797	CCCGCTTTACTGTATGATGCTGCTGTTGGCTCCCGAATCTAAGCGCGCAGCTTCGGGGG	856
QY	1036	CCGCGAGATGCGAAGGGGATGCGCAAAAGCTCCGTGCTCTTGTGAGCGCCACACCC	1095

Db	857	CCGCGGATGGCGAAGGGGGGANTG-CGCAAAAGCTCGTGCTCTTTGTGAGACGCACACACC	915
QY	1096	TACCCCTCCAGGGGGGTAGCCACTC--GCCCGCGCCCAAGAGGAGGACCACTGCCAGGCATTTCAC	1153
Db	916	TACCCCTCCAGGGGGGTAGCCACTCGCGCCCGCCCAAGAGGAGGACCACTGCCAGGCATTTCAC	975
QY	1154	CCAGCTGGTGCCTCCAAAGACAGAGCCGGGGGCGAGCGGACACTCCCAAAATTATCTATG	1213
Db	976	CCAGCTGGTGCCTCCAAAGACAGAGCCGGGGGCGAGCGGACACTCCCAAAATTATCTATD	1035
QY	1214	GGTGGGTTTACACTCTTCACGGCCCCCTGTGAAGCGGGGACTTTTCCATTCCCTCAACCCAC	1270
Db	1036	GGTGGGTTTACACTCTTCACGGCCCCCTGTGAAGSGGGAC-TTTCATTTCCTCAACCCAC	1091
RESULT 15			
BX339727			
LOCUS		1038 bp	mRNA
DEFINITION	BX339727 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA		linear EST 07-APR-2004
ACCESSION	BX339727		
VERSION	BX339727.2	GI:46271262	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1038)		
AUTHORS	Li,W.B., Gruber,C., Jesssee,J. and Polayes,D.		
TITLE	Full-length CDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	On May 2, 2003 this sequence version replaced gi:30335832.		
	Contig: 200000000		

## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

This Page Blank (uspto)

BEST AVAILABLE COPY

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 28, 2005, 05:53:55 ; Search time 80 Seconds  
(without alignments)  
1300.482 Million cell updates/sec

Title: US-10-031-589-4  
Perfect score: 1437  
Sequence: 1 MKTPELVASGMQTLGLQHR.....SVSPRRRGATQKRPWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues  
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1437	100.0	269	4 AAB67330	AAB67330 Human neu
2	1437	100.0	269	4 AAB94540	AAB94540 Human pro
3	1216	84.6	229	8 AD057487	AD057487 Hairless
4	863	60.1	177	4 AAB94561	AAB94561 Human pro
5	239	16.6	467	4 ABB11341	ABB11341 Human typ
6	218	15.2	65	5 ABB35240	ABB35240 Human ORF
7	147.5	10.3	1706	7 ABO83014	ABO83014 Pseudomon
8	136.5	9.5	4019	4 AAE13839	AAE13839 Human lun
9	136.5	9.5	4019	7 ADB6733	ADB6733 Human lun
10	136.5	9.5	4019	7 ADE87987	ADE87987 Human lun
11	136.5	9.5	4025	5 ABB69736	ABB69736 Human pol
12	136	9.5	2703	4 ABB63299	ABB63299 Drosophi
13	136	9.5	2703	5 ABB63299	ABB63299 Larval vi
14	135.5	9.4	920	7 ABO80172	ABO80172 Pseudomon
15	134.5	9.4	19938	6 ABB76681	ABB76681 Streptomy
16	134	9.3	502	6 ADA24266	ADA24266 Human WAS
17	132.5	9.2	470	4 ABB21932	ABB21932 Novel hum
18	132.5	9.2	488	7 ADG73027	ADG73027 Pseudomon
19	132.5	9.2	488	7 ADL12082	ADL12082 Pseudomon
20	132.5	9.2	580	8 AD019877	AD019877 Human sof
21	132.5	9.2	1008	7 ABO80673	ABO80673 Pseudomon
22	131.5	9.2	705	4 AAM39328	AAM39328 Human pol
23	131.5	9.2	714	4 AAM41114	AAM41114 Human pol
24	129.5	9.0	791	5 ABB65089	ABB65089 Hypoxia-r
25	129.5	9.0	791	8 ADN03640	ADN03640 Antipsoi

26	129.5	9.0	791	8 ADP22972	ADP22972 PRO polyp
27	129.5	9.0	863	6 ABB98856	ABB98856 Human sex
28	129.5	9.0	863	7 ADC31071	ADC31071 Human nov
29	129.5	9.0	863	7 ADE48296	ADE48296 Human MIC
30	129.5	9.0	863	7 ADE48308	ADE48308 Human MIC
31	129.5	9.0	863	8 ABB82355	ABB82355 Tumour-as
32	129	9.0	934	8 ADN38399	ADN38399 Novel hum
33	128.5	8.9	501	7 ABO68984	ABO68984 Pseudomon
34	128.5	8.9	536	2 ABO68008	ABO68008 Pseudomon
35	128	8.9	504	2 AAY24091	AAY24091 Human MIs
36	128	8.9	566	6 ADA24286	ADA24286 Protein r
37	127.5	8.9	1870	7 ADJ68314	ADJ68314 Human hea
38	126.5	8.8	693	7 ABO77325	ABO77325 Pseudomon
39	126.5	8.8	1199	7 ADD46005	ADD46005 Rat Prote
40	126.5	8.8	1199	7 ADE61171	ADE61171 Rat Prote
41	126.5	8.8	1321	8 ADN98332	ADN98332 Mouse Ins
42	126	8.8	718	5 ABB79534	ABB79534 Human tum
43	126	8.8	958	6 ABB75918	ABB75918 Human sec
44	125.5	8.7	641	7 ADG48274	ADG48274 Human ret
45	125.5	8.7	663	7 ADG48269	ADG48269 Human ret

## ALIGNMENTS

RESULT 1  
ID AAB67330 standard; protein; 269 AA.  
XX  
AC AAB67330;  
XX  
DT 23-APR-2001 (first entry)  
XX  
DE Human neuron progenitor cell clone #2 protein.  
XX  
KW Neuron; progenitor cell; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200107607-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 21-JUL-2000; 2000MO-JP004895.  
XX  
PR 23-JUL-1999; 99JP-00209817.  
XX  
PR 18-OCT-1999; 99US-0159528P.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
Ota T, Isogai T, Nishikawa T, Kawai Y;  
WPI, 2001-182791/18.  
XX  
PT New human polynucleotides, particularly DNAs, isolated from a cDNA  
PT library derived from progenitor cells, useful in gene therapy, as well as  
PT in producing proteins useful as diagnostic markers in drug development.  
XX  
PS Claim 1; Page 36-37; 54pp; English.  
XX  
CC The present invention relates to human proteins isolated from clones from  
CC neuron progenitor cells. The proteins and the DNA encoding them may be  
CC used in the preparation of treatments for diseases associated with the  
CC proteins  
XX  
SQ Sequence 269 AA;  
Query Match 100.0%; Score 1437; DB 4; Length 269;  
Best Local Similarity 100.0%; Pred. No. 2,7e-114;  
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MKTPELVASGMQTLGLQHRGGRVAKRTSYDETLFGSPAGTRPPPPDPWVEKA 60  
|||||

Db 1 MKTPVELAVSGMOTLGIQHRGCGYRVKARTSYVDLTFGSPAGTRPTPPDFPPWVEKA 60  
 QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTFRKKNKRPISHTPSYCDSEIFGSRSE 120  
 Db 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTFRKKNKRPISHTPSYCDSEIFGSRSE 120  
 QY 121 GASFGAPRMAKGDAAKLRALMLTPTPPRGSHSPRPREAPLRAIHAPGSKTEPGPAADS 180  
 Db 121 GASFGAPRMAKGDAAKLRALMLTPTPPRGSHSPRPREAPLRAIHAPGSKTEPGPAADS 180  
 QY 181 OKLSMGGHSSRPLKRGLSHLTLNVPSTGHPATSAFHTNGPDLRPSISGVTFRSPLV 240  
 Db 181 OKLSMGGHSSRPLKRGLSHLTLNVPSTGHPATSAFHTNGPDLRPSISGVTFRSPLV 240  
 QY 241 TSRAVSISVSPSTRRGATOKPKPPWK 269  
 Db 241 TSRAVSISVSPSTRRGATOKPKPPWK 269

# RESULT 2

AAB94540

ID AAB94540 standard; protein; 269 AA.

AC AAB94540;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:15284.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ihli S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 15284; 2537pp + Sequence listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dT primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX Sequence 269 AA;

Query Match 100.0%; Score 1437; DB 4; Length 269;

Best Local Similarity 100.0%; Pred. No. 2.7e-114;

Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSGMOTLGIQHRGCGYRVKARTSYVDLTFGSPAGTRPTPPDFPPWVEKA 60

Db 1 MKTPVELAVSGMOTLGIQHRGCGYRVKARTSYVDLTFGSPAGTRPTPPDFPPWVEKA 60

QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTFRKKNKRPISHTPSYCDSEIFGSRSE 120

Db 61 NRTGVGKEASKALGAKGSCETTPSRGSTPTLTFRKKNKRPISHTPSYCDSEIFGSRSE 120

QY 121 GASFGAPRMAKGDAAKLRALMLTPTPPRGSHSPRPREAPLRAIHAPGSKTEPGPAADS 180

Db 121 GASFGAPRMAKGDAAKLRALMLTPTPPRGSHSPRPREAPLRAIHAPGSKTEPGPAADS 180

QY 181 OKLSMGGHSSRPLKRGLSHLTLNVPSTGHPATSAFHTNGPDLRPSISGVTFRSPLV 240

Db 181 OKLSMGGHSSRPLKRGLSHLTLNVPSTGHPATSAFHTNGPDLRPSISGVTFRSPLV 240

QY 241 TSRAVSISVSPSTRRGATOKPKPPWK 269

Db 241 TSRAVSISVSPSTRRGATOKPKPPWK 269

RESULT 3

AD057487

ID AD057487 standard; protein; 229 AA.

AC AD057487;

DT 29-JUL-2004 (first entry)

DE Hairless protein interaction partner #34.

KW human; hairless protein; Hrt; ubiquitous receptor UR; MAP1A;

KW KIAA0930 protein; monocytes antigen CD14; sphingolipid activator protein; 2;

KW beta-synuclein; C11 protein; vesicle-associated membrane protein 2;

KW aldolase A; CGI-106 protein; hypothalamus protein HSMNP1; alpha enolase;

KW POM-2P3; guanine oxidoreductase; pumilio 1; VPS41; KIAA0614 protein;

KW splicing factor CCl.4; ubiquitin; beta-mannosidase; hair growth.

OS Homo sapiens.

PN US2004086945-A1.

PD 06-MAY-2004.

PR 02-JUN-2003; 2003US-00452858.

PR 03-JUN-2002; 2002US-0385414P.

XX (PROC ) PROCTER & GAMBLE CO.

XX Sreekrishna K, Gerwe GS, Toerner DR;

XX WPI; 2004-430095/40.

XX N-FSDB; AD057486.

XX New composition comprising mouse truncated hairless protein-human

XX interacting partner protein or nucleic acid complexes, useful for

XX screening test compounds that inhibit or enhance hair growth.



XX Example 1; SEQ ID NO 79; 60pp; English.  
 CC The invention relates to a composition comprising a mouse truncated  
 CC hairless (Hr) protein-human interacting partner protein or nucleic acid  
 CC complex. The human interacting partner protein comprises a molecule  
 CC selected from ubiquitin, receptor UR, MAP1A, KIAA0930 protein, monocytes  
 CC antigen CD14, sphingolipid activator protein, beta-synuclein, C11  
 CC protein, vesicle-associated membrane protein 2, aldolase A, CGI-106  
 CC protein, hypoxanthine protein HSNP1, alpha enolase, POM-2P3, quinone  
 CC oxidoreductase, pumilio 1, VPS41, KIAA0614 protein, splicing factor  
 CC CCL1.4, ubiquitin, beta-mannosidase. Also described are: methods of  
 CC assaying a test compound for agonist or antagonist activity for the above  
 CC composition; and methods of inhibiting or increasing hair growth on a  
 CC surface in a subject. The present sequence represents a hairless protein  
 CC human interacting partner used in the composition.

Query Match 84.6%; Score 1216; DB 8; Length 229;  
 Best Local Similarity 98.7%; Pred. No. 1.6e-95;  
 Matches 226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 11 GMDTGLGHRGCGYRVARTSYVDETLFGSPAGTRPPPPDPMPVWEKANTRGVKEA 70  
 Db 1 GMDTGLGHRGCGYRVARTSYVDETLFGSPAGTRPPPPDPMPVWEKANTRGVKEA 60  
 QY 71 SKALGAKSCETTPRGSTPTLTPRKKNKYPRIISHTPSYCDLSLGSRSSEGSFGAPRMA 130  
 Db 61 SKALGAKSCETTPRGSTPTLTPRKKNKYPRIISHTPSYCDLSLGSRSSEGSFGAPRMA 120  
 QY 131 KGDAKALRALMTPTTPRGSHSPRPREAPLAIHPAGSKTEPGPADSGLSMGGLHS 190  
 Db 121 KGDAKALRALMTPTTPRGSHSPRPREAPLAIHPAGSKTEPGPADSGLSMGGLHS 180  
 QY 191 SRPLRGKLSHSLTLNVPSTGHPATSPHTNGPDLRPESTGVTFRSP 239  
 Db 181 SRPLRGKLSHSLTLNVPSTGHPATSPHTNGPDLRPESTGVTFRSP 229

RESULT 4  
 AAB94561 standard; protein; 177 AA.

XX AC AAB94561;  
 XX DT 26-JUN-2001 (first entry)  
 XX DE Human protein sequence SEQ ID NO:15335.  
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX OS Homo sapiens.  
 XX PN EP1074617-A2.  
 XX PD 07-FEB-2001.  
 XX PF 28-JUL-2000; 2000EP-00116126.  
 XX PR 29-JUL-1999; 99JP-00248036.  
 XX PR 27-AUG-1999; 99JP-00300253.  
 XX PR 11-JAN-2000; 2000JP-00118776.  
 XX PR 02-MAY-2000; 2000JP-00183767.  
 XX PR 09-JUN-2000; 2000JP-00241899.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Iwagai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.

XX Claim 8; SEQ ID NO 15335; 2537bp + Sequence Listing; English.  
 XX The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo- or primer and an oligonucleotide which comprises one of the 5602  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification; where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

Sequence 177 AA;  
 Query Match 60.1%; Score 863; DB 4; Length 177;  
 Best Local Similarity 99.4%; Pred. No. 1.5e-65;  
 Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPELVSGMDTGLGHRGCGYRVARTSYVDETLFGSPAGTRPPPPDPMPVWEA 60  
 Db 1 MKTPELVSGMDTGLGHRGCGYRVARTSYVDETLFGSPAGTRPPPPDPMPVWEA 60  
 QY 61 NTRGVGEASKALGAKSCETTPRGSTPTLTPRKKNKYPRIISHTPSYCDLSLGSRSSE 120  
 Db 61 NTRGVGEASKALGAKSCETTPRGSTPTLTPRKKNKYPRIISHTPSYCDLSLGSRSSE 120  
 QY 121 GASFGAPPMAGDAKALRALMTPTTPRGSHSPRPREAPL 161  
 Db 121 GASFGAPPMAGDAKALRALMTPTTPRGSHSPRPREAPV 161

RESULT 5  
 AAB11341 standard; peptide; 467 AA.

XX AC AAB11341;  
 XX DT 11-JAN-2002 (first entry)  
 XX DE Human type II procollagen homologue, SEQ ID NO:1711.  
 XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; actinin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; anti-inflammatory;  
 KW antiaesthetic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cyostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnerary; antitumor.

OS Homo sapiens.  
 XX WO200157188-A2.  
 XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US003800.  
 XX 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Dermanac RT;  
 PI WPI, 2001-457740/49.  
 DR N-PSDB; ABA08585.  
 XX Human proteins and DNA encoding sequences useful for preventing, treating  
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis  
 PT and cancer.  
 PS Claim 20; Page 167-168; 1963pp; English.  
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC hematopoietic regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention  
 XX  
 SQ Sequence 467 AA;  
 Query Match 16.6%; Score 239; DB 4; Length 467;  
 Best Local Similarity 29.4%; Pred. No. 8.9e-12;  
 Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;  
 QY 5 VELAVSMQTLGQHR--CGGYR-----VKARTSVVDLTGSPRGCTPTPPDPDP 56  
 DB 176 WLCTLAQQAQAPGAPLPHRTSIRPGMKRLTEPAMAR--HRRPWGQRCAGVAPQGAAPP 232  
 QY 57 VEKANRTGVGKESKALGAKGSCETTPSRGSTPTLTLPKKNKRPISHPSYC---DES 113

DB 233 SHQGRT-----NTDPS--ATPRLT-----VMSGCLAPDLK 261  
 QY 114 LFGSRSEAGSFGAPRNAGDAKLRLALWTPPTPGSGSPRPREAPLPAIHAGSKTE 173  
 DB 262 APASGPRGMRGMQGS-----GALLWTPPTPGSGSPRPREAPLPAIHAGSKSR 314  
 QY 174 RGPAAOSQKLSMGGLHSSRLKRG-----LSHSL-----THLVNPGTHPATS 216  
 DB 315 AGASGRLEPVTYGMVTLFTPEAGTFLIPSPYMSPLVYQPPVPTQGLRISGLPRQG 374  
 QY 217 APTTNGPQDLRPSTSGVTF-----RSPLYTSRASVSISVSTPRR----- 257  
 DB 375 XP-SGAPFWLI-PQLAQLAQCHLPHDEVGPRRQSLGND---TLSSGLPMGRQVWPL 429  
 QY 258 ---GG--ATQKP---KPPW 268  
 DB 430 ARVGHSSPREPQVLKKPLM 449  
 RESULT 6  
 ID ABB35240  
 ABP35240 standard; protein; 65 AA.  
 AC ABB35240;  
 XX 08-JUL-2002 (first entry)  
 DT Human ORP4213 protein, SEQ ID NO:8426.  
 XX  
 KW Human; ORF; open reading frame; ORF; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotropic; antipruritic; antidiabetic; cyostatic; noctropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; immunomodulator;  
 KW cardiant; hypotensive; antichryloid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 OS Homo sapiens.  
 XX WO200190366-A2.  
 XX 29-NOV-2001.  
 XX 24-MAY-2001; 2001WO-US017076.  
 XX 24-MAY-2000; 2000US-020690P.  
 XX (CURA-) CURAGEN CORP.  
 XX Leach MD, Shinkets RA;  
 XX WPI, 2002-106200/14.  
 DR N-PSDB; ABB79266.  
 DR Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.  
 PS Claim 10; Page 2350; 2508pp; English.  
 XX Sequences ABB1028-ABB35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABB75054-  
 CC ABB79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to

the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polynucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polynucleotides and polypeptides, methods of screening for modulators of ORFX expression or activity, and methods of screening individuals for a predisposition to an ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokine, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, tissue growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, anti-inflammatory activity, tumour inhibition activity, and anti-infective activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol, ester storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases

Query Match	15.2%	Score 218	DB 5	Length 65
Best Local Similarity	71.4%	Pred. No. 5e-11		
Matches 40; Conservative	4	Mismatches 12	Indels 0	Gaps 0

```

QY      12 MQTGLQHRCRGGYRVKARTSYVDETLGSPAGTRTPDPDFPWEVEKARTRGVG 67
      | : : ||||| ||||| ||||| ||||| : |||||
Db      1 MKAHLDQRSPTSYRVKARASYVDETLGSPARTRAQPDFDPWVQNCNRSRGVG 56

```

RESULT 7	
ABO83014	
ID ABO83014	standard; protein; 1706 AA.

DT 29-JUL-2004 (first entry)

**Pseudomonas aeruginosa** polypeptide #15189.

KW Bacterial infection; *Pseudomonas aeruginosa* infection; antibacterial.

**OS** *Pseudomonas aeruginosa*.

PN US6551795-B1.

PD 22-APR-2003

PF 18-FEB-1999; 99US-00252991.

PR 18-FEB-1998; 98US-0074788P.

XX  
XX

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

DR WPI; 2003-615309/58.

DR N-PSDB; ABD16585.  
yy

PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of

PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 31760; 455bp; English.  
nu

CC The invention relates to pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABO67826-  
CC ABO84936 represent P. aeruginosa polypeptides of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
CX  
CX Sequence 1706 Aa;

Query Match	10.3%	Score 147.5	DB 7	Length 1706
Best Local Similarity	24.7%	Pred. No. 0.0028		
Matches 64; Conservative	21	Mismatches 103	Indels 71	Gaps 9

Oy	19	HHCRCGRVAKR	SYVDEL	FGSPAGTR	PTPPED	PPWEKANTRG	VGCEASAL	GAKG	78
Db	1277	HARRASLHALART	-----	GLTAGRA	TTGEBDKR	ILQAPR	PKYR	-----	1314

Qy	79	SCETT	P	S	G	S	T	P	L	T	R	K	N	K	T	R	P	I	--	S	H	T	S	V	C	D	E	S	L	E	G	S	R	S	G	S	A	F	G	A	P	R	M	A	K	G	D	A	A	K	136
Db	1315	R	A	H	N	P	R	G	S	I	T	R	A	T	P	P	R	E	G	K	O	R	--	G	S	P	A	N	P	A	S	A	G	R	R	-----	136														

QY 137 LEALLWPTPTGSHSP-----RPREALRAIHDPAGPSTKEPCGAADSOQLSMWGLHSSR 192  
 Db 1365 -----PPRRRGSPARPFCEPRQRPRRGTOFQSPARTPAED-----140C-----

Oy 193 PLKRGSLSTLHLNVSTGHPATSSAPHNNGPOLRSPSTGVTRSPLV-----TSRARSV 247  
 Db 1403 --RNNHHPGVTARPPATGRPMPTATPGAHPRHRRTATAARRCGRPSVPRATGTRRSR-- 1455

```
Qy      248 SISVSPTRRGATQKPKP 266
          : | : | : |
Db      1459 --TAPGAPQPAARQPPRP 1475
```

RESULT 8  
AAE13839

AC AAE13839;

DT 26-FEB-2002 (first entry)

DE Human lung tumour-specific protein SCC2-29.

KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;

XX

XX

XX

XX

XX

PR 29-MAR-2000; 2000US-00538037.

PR 18-AUG-2000; 2000US-00640878.

PR 01-NOV-2000; 2000US-00704512.  
PR 14-DEC-2000; 2000US-00738973.  
XX (CORI-) CORIXA CORP.  
XX  
XX  
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
XX WPI; 2001-639201/73.  
XX N-PSDB; AAD33451.  
XX  
PT New human lung-specific polynucleotides and polypeptides for the  
PT diagnosis and treatment of disease e.g. lung cancer.  
XX  
XX Disclosure; Page 309-318; 378pp; English.  
XX  
XX The invention relates to isolated lung tumour-specific proteins and their  
XX corresponding cDNA molecules. Lung tumour-specific proteins and their  
XX antigen-presenting cells are useful for stimulating and/or expanding T  
XX cells specific for a tumour protein, and for inhibiting the development  
XX of cancer. The invention also relates to a composition useful for  
XX stimulating an immune response, and for treating cancer. The lung tumour  
XX specific oligonucleotide is useful in gene therapy and for diagnosis,  
XX detection and treatment of lung cancer. The present sequence is human  
XX lung tumour-specific protein  
SQ Sequence 4019 AA;  
Query Match 9.5%; Score 136.5; DB 4; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.069;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;  
QY 29 ARTSYDETLFGSPAGR-----PPPPDDPPVWEK-----ARTRGVCK 68  
DB 988 SRPLQNNETTANRPSPVRLDCSSSTTNDPYAKPDPTRPVMTDQPKSIGLSRSPVSE 1047  
QY 69 EASKALGAKGSC-----ETTPSRGSTPTLTPRKKNK-----YRPISHTPS 108  
DB 1048 QTKAGPPIAAGTSDHFTKPSRADVPQQRIPDSYARLILTPAPLDSGPGFKTPMOPPS 1107  
QY 109 YCDESLFGSRSEGASFGAPPMAKDAKLRALIMTPTPTPGSH--SPRPREAPLRAIHP 166  
DB 1108 SQDP--YGSVSQ-----ASRLTSDPYERPAL--TPRPIDNFHNSQNDPYSQPLTPHP 1158  
QY 167 -----AGPSK--TEBG-----PADSQTLSMGGLHS-----SRPL 194  
DB 1159 AVNESFAPHRASQPTISRPTSQDPYSGPPGTPRPVVDSYSSSGTARASNTDPSQP- 1217  
QY 195 KRGLSHSLTLNVPSTGHPTASAPHTNGPOLDRPSTSGVTFRSPPLVTSRARSVSISVPST 254  
DB 1218 -----PGTRPPTTVDPYSQQPQTPRPSTGTDLFVTPVTVNQHSDDPYAHHPGT 1264  
QY 255 PRRGATQKPKPP 267  
DB 1265 PRPGISVPSQPP 1277  
RESULT 9  
ADD66733  
ID ADD66733 standard; protein; 4019 AA.  
XX  
XX ADD66733;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
XX Human lung tumour-specific related protein, SEQ ID No 425.  
XX  
XX expression control; cancer; T cell; tumour; immune; cytosolic; vaccine;  
XX human; lung tumour-specific.  
XX Homo sapiens.  
XX  
XX  
PN WO200292001-A2.

XX  
PD 21-NOV-2002.  
XX  
XX 10-MAY-2002; 2002WO-US014975.  
XX  
XX 11-MAY-2001; 2001US-00854133.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Lodes MJ, Wang T, Fan L, Algate PA, McNeill PD;  
XX WPI; 2003-120592/11.  
XX  
PT New polynucleotide and polypeptide, useful for preparing a composition  
PT for diagnosing, treating or preventing cancer.  
XX  
XX Disclosure; SEQ ID NO 425; 494pp; English.  
XX  
XX The invention relates to a novel isolated polynucleotide comprising one  
XX of 32 47-6080 base pair sequences, given in the specification, or their  
XX complements or degenerate variants, at least 20 contiguous residues of a  
XX sequence in, or having at least 75 or 90 % identity with the isolated  
XX polynucleotide, or that hybridise with the polynucleotide. The invention  
XX further comprises: an isolated polypeptide; an expression vector  
XX comprising the polynucleotide operably linked to an expression control  
XX sequence; a host cell transformed or transfected with the expression  
XX vector; an isolated antibody or its antigen-binding fragment that  
XX specifically binds to the polypeptide; a method for detecting the  
XX presence of a cancer in a patient; a fusion protein comprising the  
XX polypeptide; an oligonucleotide that hybridises to the isolated  
XX polynucleotide under moderately stringent conditions; a method for  
XX stimulating and/or expanding T cells specific for a tumour protein; an  
XX isolated T cell population; a composition comprising a first component  
XX consisting of carriers and immunostimulants and a second component; a  
XX method for stimulating an immune response in a patient; a method for  
XX treating cancer in a patient; a method for determining cancer in a  
XX patient; a diagnostic kit comprising at least one oligonucleotide or  
XX antibody and a detection reagent comprising a reporter group; and a  
XX method for inhibiting the development of cancer in a patient. The  
XX compositions of the invention have cytostatic activity and can be used to  
XX create a vaccine. The isolated polynucleotide is useful for preparing a  
XX composition for diagnosing, treating or preventing cancer. This sequence  
XX represents a human lung tumour-specific protein relating to the  
XX invention.  
SQ Sequence 4019 AA;  
Query Match 9.5%; Score 136.5; DB 7; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.069;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;  
QY 29 ARTSYDETLFGSPAGR-----PPPPDDPPVWEK-----ARTRGVCK 68  
DB 988 SRPLQNNETTANRPSPVRLDCSSSTTNDPYAKPDPTRPVMTDQPKSIGLSRSPVSE 1047  
QY 69 EASKALGAKGSC-----ETTPSRGSTPTLTPRKKNK-----YRPISHTPS 108  
DB 1048 QTKAGPPIAAGTSDHFTKPSRADVPQQRIPDSYARLILTPAPLDSGPGFKTPMOPPS 1107  
QY 109 YCDESLFGSRSEGASFGAPPMAKDAKLRALIMTPTPTPGSH--SPRPREAPLRAIHP 166  
DB 1108 SQDP--YGSVSQ-----ASRLTSDPYERPAL--TPRPIDNFHNSQNDPYSQPLTPHP 1158  
QY 167 -----AGPSK--TEBG-----PADSQTLSMGGLHS-----SRPL 194  
DB 1159 AVNESFAPHRASQPTISRPTSQDPYSGPPGTPRPVVDSYSSSGTARASNTDPSQP- 1217  
QY 195 KRGLSHSLTLNVPSTGHPTASAPHTNGPOLDRPSTSGVTFRSPPLVTSRARSVSISVPST 254  
DB 1218 -----PGTRPPTTVDPYSQQPQTPRPSTGTDLFVTPVTVNQHSDDPYAHHPGT 1264  
QY 255 PRRGATQKPKPP 267

Db 1265 PRPGISVPSQPP 1277

RESULT 10

ID ADE87987 standard; protein; 4019 AA.

AC ADE87987;

DT 29-JAN-2004 (first entry)

DE Human lung tumour antigen polypeptide #92.

KW Human; lung tumour antigen; cancer; lung cancer; CD4+; CD8+; T cell; immune response; immunostimulant; cytostatic.

OS Homo sapiens.

PN US2003118599-A1.

PD 26-JUN-2003.

PF 10-MAY-2002; 2002US-00144649.

PR 02-APR-1999; 99US-00285323.

PR 09-AUG-1999; 99US-00370838.

PR 30-DEC-1999; 99US-00476235.

PR 03-MAR-2000; 2000US-00518809.

PR 29-MAR-2000; 2000US-00538037.

PR 05-JUN-2000; 2000US-00588937.

PR 18-AUG-2000; 2000US-00640878.

PR 20-SEP-2000; 2000US-00667170.

PR 01-NOV-2000; 2000US-00704512.

PR 14-DEC-2000; 2000US-00738973.

PR 11-MAY-2001; 2001US-00854133.

PA (CORI-) CORIXA CORP.

PI Algate PA, Lodes MJ, Wang T, Fan L, McNeill PD;

XX WPI; 2003-897103/82.

DR N-PSDB; ADE87984.

XX New polynucleotides encode lung tumor antigens and are useful to

PT stimulate an immune response or detect or treat a cancer in a patient,

PT particularly lung cancer.

XX

XX Disclosure; SEQ ID NO 425; 63pp; English.

PS

XX The invention relates to polynucleotides encoding lung tumour antigens.

CC The invention also relates to the polypeptides encoded by the

CC polynucleotides, isolated antibodies or antigen-binding fragments that

CC specifically bind the polypeptides and a method for detecting cancer in a

CC patient, comprising obtaining a biological sample from the patient,

CC contacting the sample with a binding agent that binds a polypeptide of

CC the invention, detecting in the sample an amount of polypeptide that

CC binds to the binding agent, and comparing the amount of polypeptide to a

CC predetermined cut-off value. T cells specific for a tumour protein can be

CC stimulated and/or expanded by contacting the T cells with a polypeptide,

CC polynucleotide or an antigen-presenting cell that expresses a

CC polypeptide. Cancer development can be inhibited by incubating CD4+

CC and/or CD8+ T cells isolated from a patient with a polypeptide,

CC polynucleotide or an antigen-presenting cell that expresses a

CC polypeptide, so that the T cells proliferate. The invention is used to

CC stimulate an immune response or to detect or treat a cancer in a patient,

CC particularly lung cancer. This sequence represents a human lung tumour

CC antigen polypeptide of the invention. Note: The sequence data for this

CC patent did not form part of the printed specification but was obtained in

CC electronic format from USPTO at [seqdata.uspto.gov/Sequence.html](http://seqdata.uspto.gov/Sequence.html).

XX

XX Sequence 4019 AA;

XX

XX Query Match

XX 9.5%; Score 136.5; DB 7; Length 4019;

Best Local Similarity 24.0%; Pred. No. 0.069; Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 26 ARTSYDETLGSPGACTR-----PTPDFDPMPVYK-----ANRTGCVK 68

Db 988 SRPLQWNETTANRPSPVLDLSSSTTNDPVAKPPTDTPVWTDQFPKSLGSRSPVSE 1047

QY 69 EASKALGAKGSC-----ETTSRGSSTPLTPRKKK-----YRPSHPSS 108

Db 1048 QIAKGPPIAAGTSDHTKSPRADVFQRIIPSVARPLTPAPLDLSSGSGPFTPMQPPS 1107

QY 109 YCDESLFSGSRSGASFGAPMAKGDPAKRLALWTPPTPGSH--SPRREAPRAIHP 166

Db 1108 SQDP--YGSVSG-----ASRLSDVPYRPL--TPRIIDFNHNSQNDPYSQPLTPHP 1158

QY 167 -----APSK--TEFG-----PAASQKLSMGGLHS-----SRPL 194

Db 1159 AVNESFAHPSRAFSQPTISRPTSDPYSGPGRFPPVVDVSGSGTARSNTDPSQP- 1217

QY 195 KRGLSHSLTLNLVPSGTGPAATNGPOLRPTSGVTRSPPLVTRARASVSIIVST 254

Db 1218 -----PPTPTTVDPSYQQPTPRBSTQTDLFTVPTNGRHSDDPYAHPPGT 1264

QY 255 PRGCGATQKPKRP 267

Db 1265 PRPGISVPSQPP 1277

RESULT 11

ID ABP69736 standard; protein; 4025 AA.

ABP69736

DT 20-JAN-2003 (first entry)

DE Human polypeptide SEQ ID NO 1783.

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;

KW cell-proliferative disorder; neurodegenerative disease; bacterial;

KW Parkinson's disease; Alzheimer's disease; autoimmune disease;

KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;

KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;

KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

KW antithratic.

OS Homo sapiens.

PN WO200270539-A2.

PD 12-SEP-2002.

PF 05-MAR-2002; 2002WO-US005095.

PR 05-MAR-2001; 2001US-00799451.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

PI Weinman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2002-759812/82.

DR N-PSDB; ABZ11953.

XX New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet

PT or coagulation disorders.

XX

XX Claim 9; SEQ ID NO 1783; 1012pp + Sequence Listing; English.

XX

CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences (AB211119-  
 CC AB212066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ABP68902-ABP68949) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 4025 AA;  
 Query Match 9.5%; Score 136.5; DB 5; Length 4025;  
 Best Local Similarity 24.0%; Pred. No. 0.07; Mismatches 113; Indels 97; Gaps 13;  
 Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PRPPDFDPVWEK-----ANRTGVGK 68  
 Db 994 SRPLQNMETANRSPYRDLCSSSTNNDPYAKPPDPFRPMTDQFKSLGLSRSPVSE 1053  
 QY 69 EASKALCAKSGC-----ETTPSRGSGTTLTPKKNK-----YRPSHTPS 108  
 Db 1054 QTAAGPIAAGTSDHPTKPPADVFQRIKPSYARLILTPAPLDSGPGFKTPMQPPS 1113  
 QY 109 YCDESLGSRSEGSFGAPWAKGDAKRLALMTPTTPRGSH--SPRPEAPLRAIHP 166  
 Db 1114 SQDP--YGSYSQ-----ASRLISVDYERPAL--TPRPIDNPSNOSNDPISQPLTPHP 1164  
 QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194  
 Db 1165 AVNESFAPHSRASFQPGFISNPTSQDPYSQPPTRPRFVDSYQSSGTASNTDPSYQP- 1223  
 QY 195 KRGLSHLTLNLTNVSTGHPTASGHTNGPQDLRSTGVTTRSLVTSRRASVISVPSST 254  
 Db 1224 -----GTPRPFITVDYPSQOPQTPRPSTQTLFVTVTNQRHSDPYAHPPGT 1270  
 QY 255 PRRGATQKPKP 267  
 Db 1271 PRGTSVPSQPP 1283

RESULT 12  
 ABB63299  
 ID ABB63299 standard; protein; 2703 AA.  
 XX  
 AC ABB63299;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 16689.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US009231.  
 PF  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.

XX  
 PI Venter JC, Adams M, Li FWD, Myers EM;  
 XX  
 DR WPI, 2001-656860/75.  
 DR N-PSDB; ABLO7402.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 PS Disclosure; SEQ ID NO 16689; 21pp + Sequence Listing; English.

XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 2703 AA;  
 Query Match 9.5%; Score 136; DB 4; Length 2703;  
 Best Local Similarity 22.5%; Pred. No. 0.047;  
 Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARTSYVDETLFGSPAGTRPPDPF-----DPPWEKAKNRTRGVG 67  
 Db 633 GGYKMGPGQSPFAQGVPPQPOQYPPGNYPPRQYPRGAYATGPPPPPSQA-----GAG 668  
 QY 68 KEASKALGAAGSGETTPSRGSGTTLTPKKNKXKRPISHTPSYCDLSFGSRSGASFGAP 127  
 Db 689 GANSMBSGAQAG--GYRGK-----MNHTGQYPPQWPPSQQTVVPGAPGAMVGNH 741  
 QY 128 RMAKGAAPKALRLALMTP-----PPTPRGSHSPREAPLRAIHPAGSKTEPPAADS 180  
 Db 742 VQKRG-----TPRPVVGPPPPQSGSPRLNLYKQHLQHKGGVGSPTPPQGP 791  
 QY 181 QKLSMG--GLHSRPL--KRLSHSLTLNLTNVST-----GHATSAHPH 220  
 Db 792 QGYNGPPTGMHPMPGPRPHMGPRPHGPTVMGPTPTSPQSMLOGGQPOGQASGGPES 851  
 QY 221 NGPQDLRPTSGVTFRSLVTSRRASVISVPSST 255  
 Db 852 GGPEHI-SQDNGISSGGPTGAAGMHAIVSVTTGP 885

RESULT 13  
 ABG70019  
 ID ABG70019 standard; protein; 2703 AA.  
 XX  
 AC ABG70019;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE Larval viability associated protein #18.  
 XX  
 KW Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;  
 KW rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;  
 KW oilseed rape; soybean; vegetable crop; fruit.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200257455-A2.  
 XX  
 PD 25-JUL-2002.  
 XX  
 PF 18-JAN-2002; 2002WO-US001568.  
 PR 18-JAN-2001; 2001US-0262351P.  
 XX





XX WO200268436-A1  
PN

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 08:15:06 ; Search time 27 Seconds  
(without alignments)  
743.726 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437

Sequence: 1 MKTVELAVSMQMTGLQHR.....SVSPRRRGATQKPKPPWK 269

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTCDS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147.5	10.3	1706	US-09-252-991A-31760	Sequence 31760, A
2	136.5	9.5	4019	US-09-854-133-425	Sequence 425, App
3	135.5	9.4	920	US-09-252-991A-28918	Sequence 28918, A
4	134.5	9.4	503	US-09-599-287A-2	Sequence 2, Appli
5	132.5	9.2	1008	US-09-252-991A-29419	Sequence 29419, A
6	131.5	9.2	713	US-09-949-016-9700	Sequence 9700, Ap
7	130.5	9.1	506	US-09-949-016-11282	Sequence 11282, A
8	130.5	9.1	705	US-09-538-092-945	Sequence 945, App
9	130.5	9.1	878	US-09-556-706B-2	Sequence 2, Appli
10	130.5	9.1	878	US-09-724-418A-2	Sequence 2, Appli
11	130	9.0	622	US-09-949-016-9104	Sequence 9104, Ap
12	128.5	8.9	501	US-09-252-991A-17730	Sequence 17730, A
13	128.5	8.9	536	US-09-252-991A-16754	Sequence 16754, A
14	128	8.9	507	US-09-599-287A-24	Sequence 24, Appl
15	126.5	8.8	693	US-09-252-991A-26071	Sequence 26071, A
16	126	8.8	696	US-08-906-865-4	Sequence 4, Appli
17	126	8.8	696	US-09-129-668-4	Sequence 4, Appli
18	125.5	8.7	699	US-10-237-551-143	Sequence 143, Appl
19	125.5	8.7	699	US-10-237-551-254	Sequence 254, App
20	125	8.7	1187	US-09-949-016-6513	Sequence 6513, App
21	124.5	8.7	1034	US-09-252-991A-28921	Sequence 28921, A
22	123.5	8.6	315	US-09-252-991A-16743	Sequence 16743, A
23	123.5	8.6	390	US-09-252-991A-17829	Sequence 17829, A
24	123.5	8.6	1321	US-08-317-310A-64	Sequence 64, Appl
25	122.5	8.5	1037	US-09-252-991A-17548	Sequence 17548, A
26	122	8.5	428	US-09-252-991A-25955	Sequence 25955, A
27	122	8.5	904	US-09-976-594-615	Sequence 615, App

28	121.5	8.5	309	US-09-252-991A-25386	Sequence 25386, A
29	121.5	8.5	907	US-08-783-774-2	Sequence 2, Appli
30	121.5	8.5	907	US-09-328-559A-1	Sequence 1, Appli
31	121.5	8.5	907	PCT-US95-04611A-19	Sequence 19, Appl
32	120.5	8.4	202	US-09-252-991A-32054	Sequence 32054, A
33	120.5	8.4	481	US-09-949-016-9748	Sequence 9748, Ap
34	120.5	8.4	2142	US-09-538-092-1142	Sequence 1142, Ap
35	120	8.4	345	US-09-252-991A-18076	Sequence 18076, A
36	120	8.4	405	US-09-252-991A-27573	Sequence 27573, A
37	119.5	8.3	189	US-09-252-991A-16638	Sequence 16638, A
38	118	8.2	315	US-09-252-991A-30553	Sequence 20553, A
39	118	8.2	663	US-09-252-991A-30843	Sequence 30843, A
40	117.5	8.2	432	US-09-252-991A-30848	Sequence 30848, A
41	117.5	8.2	757	US-09-949-016-7121	Sequence 7121, Ap
42	117.5	8.2	906	US-09-252-991A-32715	Sequence 32715, A
43	117	8.1	658	US-09-328-559A-2	Sequence 2, Appli
44	116.5	8.1	783	US-09-252-991A-18035	Sequence 18035, A
45	116.5	8.1	2169	US-09-949-016-6930	Sequence 6930, Ap

#### ALIGNMENTS

```
RESULT 1
US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31760
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31760

Query Match      10.3%; Score 147.5; DB 4; Length 1706;
Best Local Similarity 24.7%; Pred. No. 0.00015;
Matches 64; Conservative 21; Mismatches 103; Indels 71; Gaps 9;

QY      19 HRCRGYVVKARTSYVDITLFGSPAGTPTPTPPDPVVEKANKRRGVGKSKALGAKG 78
DB      1277 HARRSLHALART-----GTLAGRATTEDERRRIQAPRRKTR----- 1314

QY      79 SCETPSGSGTPTLTPRKKNKYRPI--SHTPSYCDSELFGRSBCGASFGAPMAKDAK 136
DB      1315 RARHPGCGITPATPPPKCKQRPAPRROAKCRQ--GSPANASGRRR----- 1364

QY      137 LRALMTPTPPPTPGSHSP---RPREAPLRAIHAPGSKTEPGPAADSOKLSMGSLSSR 192
DB      1365 -----PPRRGSPAPRFRCPRRQRRPGQPSPARTBPAPD----- 1402

QY      193 PLKRLGISLTHLVNPSGHPATSPHTNGQDARPSISGTYFRSPVY-----TSRASV 247
DB      1403 --RNAHHHGVVAPRPATGRPAFTAGAHRAHRTATVAARRGQPSVPRATGTRSR-- 1458

QY      248 SISVSTPRRGATQKPKP 266
DB      1459 --TAGAPQPAARQPGKRP 1475

RESULT 2
US-09-854-133-425
; Sequence 425, Application US/09854133
```

Query Match	9.4%	Score 135.5;	DB 4;	Length 920;
Best Local Similarity	26.3%;	Pred. No. 0.00083;		

Query Match	9.4%;	Score 134.5;	DB 4;	Length 503;
Best Local Similarity	25.8%;	Pred. No. 0.00047;		
Matches	63;	Conservative	23;	Mismatches 107; Indels 51; Gaps 11;
Qy	42	PAGTR-PTPEDEDPWWEKANRTRGVGEKASKALGAKGSCETTPSRGSTPT-LTPRKGN	98	
Db	207	FGCGROSPGCTPTPPFGNNGTALGGGSIROSPLSSSPFENRPLPTPTPSRADDDKPP	266	
Qy	99	KYRISHTPSYCDLSLFGSRESGASFGAPNNAKGDAKLRLTLMTPTPTPSGHSPPRR-	157	
Db	267	PPPPVGNRPSTLHRRAV-----PPPPQNNKPPVPYST	297	
Qy	158	--EAPLRA-IHPAGSKTEBPAADSQKLSMGGLHSSRLKGLSHSLTHLVNPSGCH-	212	
Db	298	PRPAPRHPHLRPPPPSRPGPPPLPPS-----SSGNDFTPLRQNNLISLSSSTPPPLSPGRS	354	
Qy	213	-PATSAPHTNGQDLR--PNSGVTFRSPLVTSARASVTSVSTPTPR---RGATQCK--	263	
Db	355	GPLPPPPSRPPPPVDRPPGKSG-PLPPPPVSHNGSTRAPLPTPQLPSBSGVDSPPSG	413	
Qy	264	PKPP	267	
Db	414	PRPP	417	

```

RESULT 5
US-09-252-991A-29419
; Sequence 29419, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29419
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29419

Query Match          9.2% Score 132.5; DB 4; Length 1008;
Best Local Similarity 24.8%; Pred. No. 0.0017;
Matches 76; Conservative 25; Mismatches 141; Indels 65; Gaps 13;

QY 17 LQHRGGRGGR--VKARTSYVDETLFGSP-AGTRPTPDPPWVEKANRTGVEKASK 72
DB 377 LVRRGRGRRRPVPVAGTGHGPRRPAAGGPAAGADRT---DPQGGPARSARPRARRR 432
QY 73 ALGAGSCETTSRSTPLTRK---NKYRPISTPSYCSESLFGSSSEASGAPRM 129
DB 433 ATGAPRGTPAARAGA--ALRPPGAGAGAFRLTGRPTAAQAPRGTRGAPGRSRV 490
QY 130 AKGDAXLRALLMTPTPT--PRGSHSPRPREAPLRAIHPAGSKTEPGPAADSOKLSMG 187
DB 491 AAGLRRRROPALRRPPTQFPGSSGDPARAPRRRALLPARPELDADPARTRORRACR 550
QY 188 LHS-----RPLKRGSLSHSLTLNVPSTGH-----PATSAPTN 221
DB 551 RHGAGODLAVAGPAPGKAG--RAARHAGAGDAHQDPQLAGRGALRPBARAGPAR 608
QY 222 GPQD-LRP--STSGVTRSPVLTSPARSVSISV-PSYPRRG-----G 259
DB 609 RPAPRLRPHRRARPGADHLCPAATRRRRRAGAAVPPADPRRGAHQERHHQGRRRPRAG 668
QY 260 ATQKPKP 266
DB 669 GTASFPV 675

```

```

RESULT 6
US-09-949-016-9700
; Sequence 9700, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9700

```

```

; LENGTH: 713
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9700

Query Match          9.2% Score 131.5; DB 4; Length 713;
Best Local Similarity 25.2%; Pred. No. 0.0014;
Matches 61; Conservative 18; Mismatches 110; Indels 53; Gaps 8;

QY 40 GSPACTRPTPDPPWVEKANRTG-----VKEASKALGAKSCCTTSPRGSTP 90
DB 462 GPAPQQR-PPQGGPQPGPQPGPQPGPQPGPQPGPQPGPQPGPQPGPQPGPQ 519
QY 91 TLTPKKNKVRPISTPSYCSESLFGSSSEASGAPRMAGKADAKLALLMTPTPT 149
DB 520 TSAPO-----QASQAPPTQGGQGRKPVVAGGPAAPPAAPPSPPQROAGPQATQ 574
QY 150 ---GSHSPRPREAPLRAIHPAGSKTEPGPAADSOKLSMGHLSRPLKGLSHSL 205
DB 575 TVSGPAPPKASGAPPGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ 616
QY 206 NVPSTGHPATSPHTNGPQDLRPTSGVTRSPVLTSPARSVSISVSTPRRGATQK 265
DB 617 PVPRTGPTTQOP-----RPSGPGPAGRPKQPLQAKQPSQDVPPATAAAG----- 662
QY 266 PP 267
DB 663 PP 664

```

```

RESULT 7
US-09-949-016-11282
; Sequence 11282, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11282
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11282

Query Match          9.1% Score 130.5; DB 4; Length 506;
Best Local Similarity 25.0%; Pred. No. 0.0011;
Matches 61; Conservative 24; Mismatches 108; Indels 51; Gaps 10;

QY 42 PAGTR-PTPPDPPWVEKANRTGVEKASKALGAKSCCTTSPRGSTPT--LTPRKXN 98
DB 210 PGGPRQSPGPTPTPPFPANRGITAGGGSIRSSPSSSPFNRRPLPTPSBALDDKPP 269
QY 99 KYRPISTPSYCSESLFGSSSEASGAPRMAGKADAKLALLMTPTPTPGS-----HS 153
DB 270 PPPFVGNRPSIHREAV-----PPPPQNNKPPVPST 300
QY 154 PRPREAPLRAIHPAGSKTEPGPAADSOKLSMGHLSRPLKGLSHSLTLNVPSTGH- 212
DB 301 PRPSASQAPPPPPSPSPGPPPLPPS---SSGNDERTPLRQRLSLSSSTPLPSPGRS 357
QY 213 -PATSAPTNGPQDLR--PTSGVTRSPVLTSPARSVSISVSTPR---RGATQK--- 263

```

Db 358 GPLPPPSERPPPPVDPGRSG-PLPPPPVSRNGSTRALPATPOLPSRSGVDSPPSG 416  
QY 264 PKPP 267  
Db 417 PRPP 420

RESULT 8  
US-09-538-092-945  
; Sequence 945, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538, 092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuroPatSeqFormatler Version 0.9  
; SEQ ID NO 945  
; LENGTH: 705  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P17600  
US-09-538-092-945

Query Match 9.1%; Score 130.5; DB 4; Length 705;  
Best Local Similarity 23.5%; Pred. No. 0.0017;  
Matches 61; Conservative 24; Mismatches 108; Indels 67; Gaps 9;

QY 40 GSPAGTPTPDPPVWEKANRTRG-----VKEKSKMLGAGSCETTPSRSTP 90  
Db 454 GPPAQQRP-PPQGPPQGPQGPQGPPLQORPPQOQHLSGAPAG-SPLPQRPLSP 511  
QY 91 TLTPRKKNKRRPISHTSYCDSESLFGSRSEGASGAPRMAGDAKRLALMTPTPTPR- 149  
Db 512 TSAPO-----QPSQAAPRPQGGQGRSPVAGSGAPAPAPRPSPPQQAQGPQATRQ 566  
QY 150 ----GSHSPRRPRAPLAIHPAGSKTEPCPADSQKLSMGGLHSSPLKRGLSHSLTHL 205  
Db 567 TSVSGPAPPRKASGAPPGGQGRQGPQKPPGAPGPTROASQAG----- 608  
QY 206 NVBGTGHPATSAHTNGP-----QDLAPSTSGVTFRRSP---LVTSRARSV 247  
Db 609 PVPRTGPTTQQPPSPGPPGAPGAPKPOLAQKPSQDVPVPATAAAGPPHPLNKSQSLTN 668  
QY 248 SISVSTPRRGATQKEXP 267  
Db 669 AFNLP-----EPAPP 678

RESULT 9  
US-09-556-706B-2  
; Sequence 2, Application US/09556706B  
; Patent No. 6458364  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Jackman, Winthrop  
; TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220  
; FILE REFERENCE: 7682-050-999  
; CURRENT APPLICATION NUMBER: US/09/556, 706B  
; CURRENT FILING DATE: 2000-04-24  
; PRIOR APPLICATION NUMBER: 08/783,774  
; PRIOR FILING DATE: 1997-01-15  
; PRIOR APPLICATION NUMBER: 08/229,291

; PRIOR FILING DATE: 1994-04-18  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 878  
; TYPE: PRT  
; ORGANISM: Virus  
; FEATURE:  
; OTHER INFORMATION: gp350  
US-09-556-706B-2

Query Match 9.1%; Score 130.5; DB 4; Length 878;  
Best Local Similarity 25.4%; Pred. No. 0.0022;  
Matches 68; Conservative 26; Mismatches 112; Indels 62; Gaps 12;

QY 41 SPAGTRPT-PPDPPVWEKANRTGKASGALG-----AKSCETTPSRGSP 90  
Db 578 SPTSAVTTPPNATSPTLGKTSPTSAVTTPTPNATSPVGETSPQANATHTLGTSPPT 637  
QY 91 TLTPRKKNKRRPIS-----HTPSYCDSESLFGSRSEGAS--FGAPRMAGDAKRLA-LMT 143  
Db 638 VTSQPKNATSAVTQGHNRPSNPETLSPSTSDNSTSHMGGENITQVTPASISTHVS 697  
QY 144 PPPTPR-----GSHSPRPREA-----PLRAIHPAGP-----KTEPGP 176  
Db 698 SSPEPRPGTTSQASGPGNSSTTKPGEVAVTKGTPPQNATSPQAPGQKTAVPVTSIG 757  
QY 177 AADQKLSMGGLHSSPLKRGLSHSLTHLVNPT--GHPATSAHTNGPQDLRPSG 233  
Db 758 KANS---TTGGKHTT-----GHGARTSTPTDYGSDTTPRPYRNATYLLPSTISK 807  
QY 234 ----TFRSPLVTSRARSVSISVSTPR 256  
Db 808 LRPRMTFTSPVTTAQATVPVPTSQPR 835

RESULT 10  
US-09-724-418A-2  
; Sequence 2, Application US/09724418A  
; Patent No. 6692749  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Jackman, Winthrop  
; TITLE OF INVENTION: NON SPLICING VARIANTS OF GP 350/220  
; FILE REFERENCE: 7682-054-999  
; CURRENT APPLICATION NUMBER: US/09/724,418A  
; CURRENT FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 09/556,706  
; PRIOR FILING DATE: 2000-04-24  
; PRIOR APPLICATION NUMBER: 08/229,291  
; PRIOR FILING DATE: 1994-04-18  
; PRIOR APPLICATION NUMBER: 08/917,320  
; PRIOR FILING DATE: 1997-08-25  
; PRIOR APPLICATION NUMBER: 08/783,774  
; PRIOR FILING DATE: 1997-01-15  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 878  
; TYPE: PRT  
; ORGANISM: Epstein-Barr Virus  
US-09-724-418A-2

Query Match 9.1%; Score 130.5; DB 4; Length 878;  
Best Local Similarity 25.4%; Pred. No. 0.0022;  
Matches 68; Conservative 26; Mismatches 112; Indels 62; Gaps 12;

QY 41 SPAGTRPT-PPDPPVWEKANRTGKASGALG-----AKSCETTPSRGSP 90  
Db 578 SPTSAVTTPPNATSPTLGKTSPTSAVTTPTPNATSPVGETSPQANATHTLGTSPPT 637  
QY 91 TLTPRKKNKRRPIS-----HTPSYCDSESLFGSRSEGAS--FGAPRMAGDAKRLA-LMT 143



```
Db 638 VTSOPKNAATVTTGQHNRPSSNPETLSPSTSDNSTSHMGENTITQVTPASISTHNVST 697
; FILE REFERENCE: 107196.136
Qy 144 PEPTP-----GSHSPRPREA-----PLAIIHAGS-----KTEBP 176
; CURRENT APPLICATION NUMBER: US/09/252,991A
Db 698 SSPEPRGTTSGASGPGNSSTKPGENVVTKTPPQNAISQASGOKTAVPTVSTG 757
; CURRENT FILING DATE: 1998-02-18
Qy 177 AADSKLWGLSHSRPLKRGSLHSLTHLVNST---GHPATSAPHTNGPQDLRPSGCV 233
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
Db 758 KANS---TTGGKHTT-----GHGARTSTETPTDYGDSSTPRRYNATYTLRPSSTSK 807
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
Qy 234 -----TFRSPLVTSRARSYSISVPSTPR 256
; US-09-252-991A-17730
Db 808 LRPRWTFTSPVTTAQTAVPVPPTSPQR 835
; US-09-252-991A-17730

RESULT 11
US-09-949-016-9104
; Sequence 9104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO: 9104
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9104

Query Match 9.0%; Score 130; DB 4; Length 622;
Best Local Similarity 29.6%; Pred. No. 0.0016;
Matches 74; Conservative 27; Mismatches 81; Indels 68; Gaps 18;

Qy 67 GKEASKALGAK-GSCETTPSRGCTPTLTFRKKNKTRPISHTPSYCDLSFGSRSGASFG 125
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 19 GPLAOTGGARIGS---SPSFTSTMDLTP-PRKYNPLR-----NESI-SLIERGASGS 67
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Qy 126 AP--RMAGSDAKLALMTPTPRGSHSPR-----PREAPLAIHPAG--PSKTEP 174
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 68 TPPEELPSPASSLAPIL--PLLP-GDPSPTTLCSPFRKSNLRLANPAGRPGSKBP 123
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Qy 175 GPADS-----OKLSMGSLHSSRPLKRG-----LSHSLTHLN 206
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 124 GRAADGGGIVGAAMPDSGLPLLDPMNKLSCGGGRRTTR-VEGGLGGEEMTRHG-SFTVN 181
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Qy 207 VPSTG--HP--ATSAAPHNN-----GPODLRPSISGVTFRSPLVTSRARSYSISVPSTPR 256
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 182 KPTRMMLPNDKMGPGVSYLVRYMGCVEVLQSMRDLDFNTTQVTR-EAISLVEAVPG 240
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Qy 257 RGATQKPKP 266
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 241 AKGATRRKKP 250
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

RESULT 12
US-09-252-991A-17730
; Sequence 17730, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Query Match 8.9%; Score 128.5; DB 4; Length 501;
Best Local Similarity 24.3%; Pred. No. 0.0016;
Matches 67; Conservative 25; Mismatches 113; Indels 71; Gaps 12;

Qy 2 KTVPELAVSGMQTLGLQHRGGRGVKARTSYVDETLFGSPAG-----TRPTPDF 52
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Db 274 RTPV-----RGRGRTAPRPPR-----LAAAGPALPARRRGRGRGPGI 314
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Qy 53 DPWVEKANRTRGVGEKASKALGAKSGCTTPSRGCTPTLTFRKKNKTRPISHTPSYCD 112
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Db 315 -----PQNPQRPVK-----RATGQRQVRRPGRACAGAVR-RGRGDRPAPR----- 357
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Qy 113 SLFGSRSGASFGAPRMAKDAKRALMTPTPRGSHSPRPREAPLAIHPGPKYT 172
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Db 358 ---GTRGGOAVALRRHRRGTGA-----PDRACGGOBERPRO---RAGQPAHPARG 402
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Qy 173 EPGPADSQKLSMGSLHSSRPL-----KR-----GLSHLTHLVNSTGHATSAPHNTNG 222
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Db 403 RPPARQGLPAPRGDPHGRRLTPRGNGCAKEAMGIPGHFHF-SSSGHNGEWPGRD 461
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Qy 223 PODLRPSTGVTFRSPLVTSRARSYSISVPSTPRG 258
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

Db 462 PRSPAPAGPRARRRSGIATIALRLIDLNRGASGREG 497
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17730
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17730

RESULT 13
US-09-252-991A-16754
; Sequence 16754, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Query Match 8.9%; Score 128.5; DB 4; Length 536;
Best Local Similarity 21.9%; Pred. No. 0.0018;
Matches 87; Conservative 27; Mismatches 113; Indels 171; Gaps 19;

Qy 18 QHRC---RGGYRVARTSYVDETLFGSPAGTRP-----TPPDPDFPWVEKANRTRG 65
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 81 RHRCLPASATYRSTCKRTISAAPSP---SCAGSTPRXMRRAAATPPSCAGTDQSRKTS 137
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Qy 66 -----VGKEASKALGAKSGCTTPSRGCTPTLTFRKKNK-----YRPI-----SHTPSYC 110
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Db 138 ASGCVSNMAR---QADSCAASGRRRAKSLQCRNRNRIGTRPLMAASQASQLHPYH 194
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754
```

```

QY      111 DESIFGR-----SEGSF 124
Db      195 GGNARARRRRSCRAALSTASSMARIAPSSSSSMPPPMRRNATRTTRAGSGGCT 254
QY      125 GAPPMAKDAKALPALMTPTPTPGSGHSPPREAPRLAHPA-----167
Db      255 GGGVTGTSSASPRPA--TPAAGAGARRARPRPRPARTCRPACRCGPGSAAPRRHRRR 312
QY      168 -----GPEKTEPG-----PAADQKLSMGGLHSSRPLKGL 198
Db      313 DRRRPSGGCGGPGPTSGRVGRVAGNCARAGRPAAARRARDG-----RPVRR-- 365
QY      199 SHSLTHLVSTGTPATAP--HTNGPDLRPTSGVTFRSPVLTSPAR---SVSISVPS 253
Db      366 -RALAGCPVPAAGRPGRWSPRRKSGSP--RPPVAG--RDPSPGRALAAAGSVADSAFG 418
QY      254 T-----PRRG-----ATOKPKP 266
Db      419 TRRRAAAGCGRRRRARKRAVPRRSGPCRPATGAPPP 456

```

```

RESULT 14
US-09-599-287A-24
; Sequence 24, Application US/09599287A
; Patent No. 6635446
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Ines M. Anton
; APPLICANT: John H. Hartwig
; APPLICANT: Ralf S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-004
; CURRENT APPLICATION NUMBER: US/09/599, 287A
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/066,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translated WIP ORF No. 6635446 3
US-09-599-287A-24

```

```

Query Match      8.9%; Score 128; DB 4; Length 507;
Best Local Similarity 25.3%; Pred. No. 0.0018;
Matches 62; Conservative 23; Mismatches 108; Indels 52; Gaps 11;

QY      42 PAGTR-PTPPDFDPWVEKANRTGCGKASKALGAKSCCTTPSRGSGTPT--LTPRKN 98
Db      210 PGGRRQPSGGTPTPPFPRNGTALGGGSIROSPLSSSPFNRRPLPTTPSRALDDKPP 269
QY      99 KYRPISTPTSYCDSESLFGSRSEGASFGAPMAKGDAAKRALMLTPTPTPGSGHSPRR- 157
Db      270 PPPVGNRRPSIHREAV-----PPPPQNNKRPVVPST 300
QY      158 ---EAPLRA-IHPAGPSTKEPGPADSOKLSMGGLHSSRPLKGLSHSLTHLVNPTGHP 213
Db      301 PRPSAPHRPRLRPPPSRPPGPPPLPS--SSGNDTERLRLPQRLNLSLSSSTPLPSPGRS 357
QY      214 A---TSAPHTNGPDLR--PSTGVTFRSPVLTSPARSVSISVPTSPR---RCGATOK-- 263
Db      358 GPLPPPVSESRPPVVRPPGRSG-PLPPPPVSRNGSTRALPATPQLPSRSGVDSPPRS 416
QY      264 -PKPP 267
Db      417 GPRPP 421

```

```

RESULT 15
US-09-252-991A-26071
; Sequence 26071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26071
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26071

```

```

Query Match      8.8%; Score 126.5; DB 4; Length 693;
Best Local Similarity 26.9%; Pred. No. 0.0038;
Matches 87; Conservative 28; Mismatches 112; Indels 97; Gaps 19;

QY      16 GLOHRCR-----GGYVKATSYVDLTFGSP---AGTRPTPPDFDPWVEKANRR 64
Db      332 GLPQRAKOPAAATGAPGTAGRGRSSGADP--GQPLQAGRRLEP-----NVRP 378
QY      65 GYVKEASKLGAAGSCETTPSRGSTPTLTP---RKKNKYRPISHTPSYCDSESLFGSRSE 120
Db      379 GVALRRRRARTGCGVGNPPAR--YDILRRPGARQLHGRPGAGDLS--GDRFAPRRAR 435
QY      121 GASFGAP---EMAKD---AAKLRLMLTP-----PPTPRGSH 152
Db      436 TALGAPAAATQVAGRRDGRSALRLALQADPTHGRRRRGRRRQVPPRQSPAPPRGDR 495
QY      153 S-PPREPAP-----LRAIHGAPGKTEPPGA-----ADQKLSMGGLHSSRPL 194
Db      496 TGRPADPAATGGLPGRPSGRLRRPH--AGPG--HPGARGRPSRADHPQPLAGLDGGRPR 552
QY      195 KR-GLSHSLTHLVNPTSGHPATGAPHTNGP-----ODLRPSTSGVTFRSPVLTSPRA 244
Db      553 SRPGVQKGLCRAGIPGTAMPAT--HAGPAAASLBERGAAVLPAGLGLRRTRRNAARS 609
QY      245 RSVSISVPS---TPRRGATOKP 264
Db      610 RRGAGQGPSRGLRRRRGRARRLP 633

```

Search completed: March 28, 2005, 08:54:14  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 08:53:20 ; Search time 390 Seconds  
(without alignments)  
228.375 Million cell updates/sec

Title: US-10-031-589-4  
Perfect score: 1437  
Sequence: 1 MKTPVELAVSGMTGLQHR.....SVSPRRRGATQKPKPMK 269

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1407402 seqs, 331100923 residues  
Total number of hits satisfying chosen parameters: 1407402

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NRM\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1216	84.6	229	US-10-452-858C-79	Sequence 79, Appl
2	239	16.6	467	US-10-276-774-1711	Sequence 1711, Ap
3	218	15.2	65	US-09-864-408A-8426	Sequence 8426, Ap
4	150	10.4	668	US-10-437-963-134726	Sequence 134726,
5	139	9.7	429	US-10-437-963-114012	Sequence 114012,
6	136.5	9.5	4019	US-09-738-973-425	Sequence 425, App
7	136.5	9.5	4019	US-09-854-133-425	Sequence 425, App
8	136.5	9.5	4019	US-10-144-649A-425	Sequence 40452, A
9	135	9.4	442	US-10-425-114-40452	Sequence 40452, A
10	134.5	9.4	503	US-10-078-547-2	Sequence 2, Appli
11	134.5	9.4	19652	US-10-084-846A-7	Sequence 7, Appli
12	134	9.3	625	US-10-437-963-143626	Sequence 143626,
13	134	9.3	937	US-10-437-963-156445	Sequence 156445,

14	133	9.3	380	16	US-10-437-963-200034	Sequence 200034,
15	133	9.3	514	15	US-10-437-963-111701	Sequence 111701,
16	132.5	9.2	488	15	US-10-365-742-20	Sequence 20, Appl
17	132.5	9.2	497	16	US-10-437-963-125004	Sequence 125004,
18	132	9.2	621	16	US-10-437-963-177697	Sequence 177697,
19	131.5	9.2	277	16	US-10-437-963-177998	Sequence 177998,
20	130.5	9.1	417	16	US-10-437-963-143835	Sequence 143835,
21	130.5	9.1	878	16	US-10-722-050-2	Sequence 2, Appli
22	130	9.0	307	16	US-10-437-963-181279	Sequence 181279,
23	130	9.0	391	15	US-10-437-963-187780	Sequence 187780,
24	129.5	9.0	791	15	US-10-170-385-57	Sequence 57, Appl
25	129.5	9.0	863	15	US-10-359-012-2	Sequence 2, Appli
26	129.5	9.0	863	15	US-10-359-012-14	Sequence 14, Appl
27	129.5	9.0	19723	15	US-10-084-846A-5	Sequence 5, Appli
28	128	8.9	507	13	US-10-078-547-24	Sequence 24, Appl
29	128	8.9	514	16	US-10-437-963-195781	Sequence 195781,
30	128	8.9	797	14	US-10-156-761-10907	Sequence 10907, A
31	127.5	8.9	388	16	US-10-437-963-160033	Sequence 160033,
32	127.5	8.9	713	16	US-10-437-963-150342	Sequence 150342,
33	127.5	8.9	1870	16	US-10-408-765A-120	Sequence 120, App
34	127	8.8	238	16	US-10-437-963-145797	Sequence 145797,
35	127	8.8	481	16	US-10-437-963-176449	Sequence 176449,
36	126.5	8.8	1321	15	US-10-694-874-4	Sequence 4, Appli
37	126	8.8	696	9	US-09-129-668-4	Sequence 4, Appli
38	126	8.8	696	14	US-10-122-805-4	Sequence 4, Appli
39	125.5	8.7	699	14	US-10-121-888-143	Sequence 143, App
40	125.5	8.7	699	14	US-10-200-562-143	Sequence 143, App
41	125.5	8.7	699	14	US-10-237-551-143	Sequence 143, App
42	125.5	8.7	699	14	US-10-237-551-254	Sequence 254, App
43	125.5	8.7	1103	16	US-10-437-963-166325	Sequence 166325,
44	125	8.7	231	16	US-10-767-701-88778	Sequence 88778, A
45	125	8.7	469	16	US-10-437-963-136216	Sequence 136216,

# ALIGNMENTS

RESULT 1  
US-10-452-858C-79  
Sequence 79, Application US/10452858C  
Publication No. US20040086945A1  
GENERAL INFORMATION:  
APPLICANT: Sreekrishna, Kotikanyadnam  
APPLICANT: Goetner, Gina S.  
TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THEREFOR  
FILE REFERENCE: 8956P  
CURRENT APPLICATION NUMBER: US/10/452,858C  
NUMBER OF SEQ ID NOS: 94  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 79  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: misc feature  
LOCATION: (217)..(217)  
OTHER INFORMATION: The 'Xaa' at location 217 stands for Lys, Arg, Thr, or Met.  
FEATURES:  
NAME/KEY: misc feature  
LOCATION: (219)..(219)  
OTHER INFORMATION: The 'Xaa' at location 219 stands for Tyr, Cys, Ser, or Phe.  
NAME/KEY: misc feature  
LOCATION: (225)..(225)  
OTHER INFORMATION: The 'Xaa' at location 225 stands for Leu, or Phe.  
US-10-452-858C-79  
Query Match 84.6%, Score 1216, DB 15, Length 229;  
Best Local Similarity 98.7%, Pred. No. 8.2e-83;  
Matches 226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



QY 186 GGLHSSRPLKRGKLSHLTLNVPSTGHPATSAHTNNGPDLRSTST-----GTFER 236  
DB 227 GGGGGYTPSPDTPPSPSSGSSPTT--PGGGGGYTPSPDTPPSPSSGSSRTTPGCGSTP 284  
QY 237 SPLVTSRRARSVISVSPSTPRRGATOKRPP 266  
DB 285 TPCGTPAPASSGTS-PTTP--GGSYYPPRP 311

## RESULT 5

US-10-437-963-114012  
; Sequence 114012, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 114012  
; LENGTH: 429  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(429)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_17745C.1.pep  
US-10-437-963-114012

Query Match 9.7%; Score 139; DB 16; Length 429;  
Best Local Similarity 24.5%; Pred. No. 0.027;  
Matches 83; Conservative 36; Mismatches 100; Indels 120; Gaps 19;

QY 21 CRGGR---YKARSYVDETLFGS-----PAG-----TRPPRP-----FDPPEVERAN 61  
DB 5 CTGVPRWPAFYRRRGRYTHROLTORSPHPPPAALVVSQPPPPDAPLAPRPPSIGAS 64  
QY 62 RTRGVGKEASKALGAKGS-----CETTPSRG-----STPPLTPRKXKX 100  
DB 65 RAKPASISAAARSRAEPAAIGCTSTPSRSPIPPNDAMPARCNRLPRPRLAVSRHAS 124  
QY 101 RPISH-----TPSYCDESLFGSRSEGASFGAPRAKGDAAKLRLALMTPTPPRSGH 152  
DB 125 NDRSHHNSRCCQMSHLAIALUTS--SAAATKRSRPPBP--QAVRCRSL---QPPAP---- 175  
QY 153 SPRPREADPLRAIHPPAGSKTEBPAPADSOKLSMGLH-----SSRPLKKGSLH- 200  
DB 176 -----AVAAARELHVAGRDSTLPPP-----HVAGLHDACHTAGLAHAEARPPRCLDHH 223  
QY 201 -----SLTHLNVSTGHPATSA--PHTNGPDLRSTSGYTRPSPL--VTSRAR----- 245  
DB 224 GRCLASQSPQMPQPPPTAAGVSHNRQPL---TGGVVAASGCIAGVTGSRXKPPPP 280  
QY 246 -----SVSISVSPSTPRRGATOKRPP 267  
DB 281 HLSGGRASVLTGGSSGSVSRKVAIAFGR-----RRKPP 315

RESULT 6  
US-09-738-973-425  
; Sequence 425, Application US/09738973  
; Patent No. US20020110563A1

; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algale, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 425  
; LENGTH: 4019  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-973-425

Query Match 9.5%; Score 136.5; DB 9; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.5;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPVAGTR-----PTPPDPPEPWEK-----ANRRGKX 68  
DB 988 SRPLQWNETTANRBPVVDLCSSSTTNDPYAKPDPTRPPVWTDQFPYSLGSRBPVSE 1047  
QY 69 EASKALGAKGSC-----ETTPSRGCTPPLTPRKXKX-----YRPISHTPS 108  
DB 1048 QTAKEPIAAGTSDHTKSPRADVFGQRIRIDSVARPLTLTAPALDSGCPFTMQPPPS 1107  
QY 109 YCDESLFGSRSEBGASFGAPRAKGDAAKLRLALMTPTPPRSGH--SPRPREADPLRAIH 166  
DB 1108 SQDP--YGSVQ-----ASRLISVDPYERPAL--TPRIDNFSHQSNDPYSQPLTPEP 1158  
QY 167 -----AGPSK--TEPG-----PADSQKXSMGSLH-----SRPL 194  
DB 1159 AVNESFAHSPRASPGTISPTSDPYSPGPTPRPVDSYSSGSGARSTNTPDYSOP- 1217  
QY 195 KRGLSHLTLNVPSTGHPATSAHTNNGPDLRSTSGVTRSPPLVTSRRARSVISVPS 254  
DB 1218 -----POTRPTTVDPYSQOPQTPRSTOTDLFTVPTVTHNRHDPVAPHPGT 1264  
QY 255 PRRGATOKRPP 267  
DB 1265 PRPGISVPSQP 1277

## RESULT 7

US-09-854-133-425  
; Sequence 425, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 425

LENGTH: 4019  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-854-133-425

Query Match 9.5%; Score 136.5; DB 9; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.5;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPPDDPPWVEK-----AARTGVCK 68  
DB 988 SRPLQNMETTNANRSPVRLDCSSSTNNDPYAKPPTPRPVMTDQPKSLGLSRSPVSE 1047  
QY 69 EASKALGAKGSC-----ETTPSRGSTRPTLRKKNK-----YRPSHTPS 108  
DB 1048 QTAKGPIAAGTSDHFTKPSRADVFQQRIPDSYARLTLPAPLDSGPGFKTPMOPPS 1107  
QY 109 YCBESLFGSRSEGSFAPPMKADAKLRALLMTPTPRGSH--SPRREAPLRAIHP 166  
DB 1108 SQDP--YGSVSO-----ASRRLSVDPYERPAL--TPRPIINFSHQNDPYSQPPLTPHP 1158  
QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194  
DB 1159 AVNESFAHPSRARSQGTISRPTSQDPYQPPGTPTPRVDSYSSQSGTARSNTDPYSQP- 1217  
QY 195 KRGLSHSLTHLVNSTGHPATSAPTNGPDLRPTSGVTFRSPLYTSRARSVISVPS 254  
DB 1218 -----PGRPTPTVDPYSQQPQTTPRSTQTDLFVTPVTHQHSDDPYAHHPGT 1264  
QY 255 PRRGATQKPKPP 267  
DB 1265 PRPGISVPSQP 1277

RESULT 8  
US-10-144-649A-425  
Sequence 425, Application US/10144649A  
Publication No. US20030118599A1  
GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Algate, Paul A.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
FILE REFERENCE: 210121.475C11  
CURRENT APPLICATION NUMBER: US/10/144,649A  
CURRENT FILING DATE: 2002-08-21  
NUMBER OF SEQ ID NOS: 749  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 425  
LENGTH: 4019  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-144-649A-425

Query Match 9.5%; Score 136.5; DB 14; Length 4019;  
Best Local Similarity 24.0%; Pred. No. 0.5;  
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPPDDPPWVEK-----AARTGVCK 68  
DB 988 SRPLQNMETTNANRSPVRLDCSSSTNNDPYAKPPTPRPVMTDQPKSLGLSRSPVSE 1047  
QY 69 EASKALGAKGSC-----ETTPSRGSTRPTLRKKNK-----YRPSHTPS 108  
DB 1048 QTAKGPIAAGTSDHFTKPSRADVFQQRIPDSYARLTLPAPLDSGPGFKTPMOPPS 1107  
QY 109 YCBESLFGSRSEGSFAPPMKADAKLRALLMTPTPRGSH--SPRREAPLRAIHP 166  
DB 1108 SQDP--YGSVSO-----ASRRLSVDPYERPAL--TPRPIINFSHQNDPYSQPPLTPHP 1158

QY 167 -----AGPSK--TEPG-----PADSQKLSMGGLHS-----SRPL 194  
DB 1159 AVNESFAHPSRARSQGTISRPTSQDPYQPPGTPTPRVDSYSSQSGTARSNTDPYSQP- 1217  
QY 195 KRGLSHSLTHLVNSTGHPATSAPTNGPDLRPTSGVTFRSPLYTSRARSVISVPS 254  
DB 1218 -----PGRPTPTVDPYSQQPQTTPRSTQTDLFVTPVTHQHSDDPYAHHPGT 1264  
QY 255 PRRGATQKPKPP 267  
DB 1265 PRPGISVPSQP 1277

RESULT 9  
US-10-425-114-40452

Sequence 40452, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovacic, David K.  
APPLICANT: Tabacka, Steven E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(5313) B  
CURRENT APPLICATION NUMBER: US/10/425, 114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 40452  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB143-063-EB\_Flt.pep  
US-10-425-114-40452

Query Match 9.4%; Score 135; DB 15; Length 442;  
Best Local Similarity 28.5%; Pred. No. 0.056;  
Matches 69; Conservative 28; Mismatches 101; Indels 44; Gaps 12;

QY 37 TLFGSPAGTRPTPPDDPPWVEKANRTGVGKASALGAKGSCETTPSRGSTRPTLR- 95  
DB 55 TTRRCRGTSPTSPPTSSAPSR-----PRPATSPPTGACRCPPTG 102  
QY 96 KKKKRPISHTSPYCDBSLFGSRG-EGASFGAPPMKADAKLRALLMTPTPRGSHSP 154  
DB 103 RRRRRPPSSAS-----ARSGAAATAAIPR-----APSPATATSPCTSPR--SP 145  
QY 155 RPREAPLRAIHPGPKTEPG--PADSQKLSMGGLHSRPLKRLSHSLTHLVNSTG 211  
DB 146 -TRTPTRRASRSPARSFGSPSPRPPTFASGISTTSRSILR--CSTISACGRCPPTG 203  
QY 212 HPATSAFH--TNGPDLRPTSGVTFRSP-----LVTSRARSVISVSTPRRGATQK 263  
DB 204 ACATWSPSTCTPRAPTPTGTRGSSAPSPASSSTASTSSRSVGAAATCASASRRSR 263  
QY 264 PK 265  
DB 264 PR 265

RESULT 10  
US-10-078-547-2

Sequence 2, Application US/10078547  
Publication No. US20020199211A1  
GENERAL INFORMATION:  
APPLICANT: Narayanasamy Rameeh  
APPLICANT: Miguel A. de la Fuente  
APPLICANT: Ines M. Anton  
APPLICANT: Ralf S. Geha  
TITLE OF INVENTION: WIP, A WASP-Associated Protein

FILE REFERENCE: 1242.1022-005  
CURRENT APPLICATION NUMBER: US/10/078,547  
CURRENT FILING DATE: 2002-02-19  
PRIOR APPLICATION NUMBER: 09/599,287  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: PCT/US98/27501  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/101,457  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/068,533  
PRIOR FILING DATE: 1997-12-23  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 2  
LENGTH: 503  
TYPE: PRT  
ORGANISM: Human  
US-10-078-547-2

Query Match 9.4%; Score 134.5; DB 13; Length 503;  
Best Local Similarity 25.8%; Pred. No. 0.071;  
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;

QY 42 PAGTR-PTPPEDPPVWEKANTRGVGEKASKALGAKGSCETTPSRGSTPT--LTTPRKKN 98  
DB 207 PGGPRQSPGPPPPPPGNGRTALGGGSIQSGPLSSSPFSNRPPLPTPSALDOKPP 266  
QY 99 KYRPISTHTPSYCDSELFSGSRSEGASFGAPRAKGAALRLMTPTPPRSGSHSPRR- 157  
DB 267 PPPPVGNPPSIHREAV-----PPPPQNNKPPVPSST 297  
QY 158 ---EAPLA-IHPAGSKTEPPGPAADSOGLSWGLHSSRPLKRGLSHLTLMNPSTGH- 212  
DB 298 PRPSAPHPHRLPPPPSPGPPPLPPS---SSGNDETPLPORKLTLSSSTPLPSGGS 354  
QY 213 -PATSAPTNGPQDLR--PSTSGVTPRSPLVTSRAVSISVPSPTP--RGATQK--- 263  
DB 355 GLPVPPEPERPPVRDPPGRSG-PLPPPPVSRVSGTSRALPLATPQLPSRSGVDSPPSG 413  
QY 264 PKPP 267  
DB 414 PRPP 417

RESULT 11  
US-10-084-846A-7  
Sequence 7, Application US/10084846A  
Publication No. US2004006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MUEHLEWEG, AGNES  
APPLICANT: TREPFER, AXEL  
APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVIARYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
PRIOR FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: PCT/EP01/09815  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO: 7  
LENGTH: 19652  
TYPE: PRT  
ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.  
US-10-084-846A-7

Query Match 9.4%; Score 134.5; DB 15; Length 19652;

Best Local Similarity 25.7%; Pred. No. 4.1;  
Matches 80; Conservative 29; Mismatches 103; Indels 99; Gaps 16;

QY 26 RYKARTSYVDETLFSSPPG-----TRPPEDPPVWEKANTRGVGEKASKAL 74  
DB 11238 RRPVPAHVDAEAAIGLVFADGVDVPSRSVPSLPPQVVER---ROSETPMSAS 11293  
QY 75 GAKGSC-----ETTPSGSTPLTP-----RKKNKTRP--- 102  
DB 11294 SSQGRCTWRSPAQSVASRWPPSPANSIDSTASRSST-LVTPCTSMTNSARRRFPSSPBA 11352  
QY 103 -ISHTPSYCDSELFSGSRSEGASFGAPRAKGAALRLMTLM-----TTP 145  
DB 11353 RIS-TPG-CPDIARVASPSAPSSSTRILAPGAANASABSGWLSRVAVAVRTPL 11410  
QY 146 PTPRGSH--SPRPAPRLAIHPAGSPKTEPPGPAADSOGLSWGLHSSRPLKRGLSHLT 203  
DB 11411 YTKSFTHGSSSTRRRRISRSFTVALP---PLSGALPSETWAGDHRGAPLSRSATYH--- 11463  
QY 204 HLNVPSSTGHPAATSAF-----HTNGPD-----LRPSTGVTRRSPLVTSRAKSV 247  
DB 11464 -----CAHSGTRSPDISISKERSAGPTAAHAARVASPSAGVSAQSP-KTSRASSV 11516  
QY 248 SISVSTPRRG 258  
DB 11517 AVATITSSRAG 11527

RESULT 12  
US-10-437-963-143626  
Sequence 143626, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 143626  
LENGTH: 625  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) ..(625)  
OTHER INFORMATION: unsure at all Xaa locations  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_44516C.1.pcp  
US-10-437-963-143626

Query Match 9.3%; Score 134; DB 16; Length 625;

Best Local Similarity 23.5%; Pred. No. 0.098;  
Matches 73; Conservative 27; Mismatches 103; Indels 108; Gaps 13;

QY 41 SPAGTRPTP-----PDPDPVWEKANTRGVGEKASKALGAKGSCETTPSRGSTPTL 92  
DB 227 APASTAXPBRPBRPPLPKSPF-REAAARPPSLDLAVSLPSGSAARPPR----- 279  
QY 93 TPRKKNKTRPISTHTPSYCDSELFSGSRSEGASFGAPRAKGAALRLMTLM----- 138  
DB 280 -----REVAFTPPRDSASASARPO-----PRGGPAAALAAATARGPOASRASP 325  
QY 139 -----ALMTPTPTPPRGSHS-----PPREAPRLAIHPAGPSK 171



```

Db      326 PRFVVRPRPAQADKGAITTPPTPMQARAGPFRVVRVPRPGRGSSPPHPLPCR 385
Qy      172 -----TERGPADDSQKLSMGSLHSRLKGLSHLNLVNP-----STGHDPATSA 217
Db      386 PGRPADCLLLSPSPSPASISLASVA---CPADSPSTLPRRTTPSSPPKSTABAPSTBA 442
Qy      218 --PHTNGPDLRPESTSGVTERSPLYTSRRAR-----SVSISV-----PST-PR 256
Db      443 SPPLJGASISLRSPSPSPPLVLTSTRQSTPSPPPPSISLASVACPADSPSTLPR 502
Qy      257 RGATQKPKPP 267
Db      503 RTRPSSPSPSP 513

RESULT 13
US-10-437-963-156445
; Sequence 156445, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 156445
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(937)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_56111C.1.pcp
US-10-437-963-156445

Query Match          9.3%; Score 134; DB 16; Length 937;
Best Local Similarity 26.2%; Pred. No. 0.15;
Matches 68; Conservative 21; Mismatches 83; Indels 88; Gaps 14;

Qy      40 GSPAG-TRTPPPDFDPWEKANRTGKVGKEASKALGAKSCETTPSRGS----- 88
Db      702 GPPPGSPSRPTOP---PEPLPSRVTR-----VDAAPTLAPSRASPPPHLSPPS 747
Qy      89 -----TPTLTPRKKKKYRPISTPSYCDSESLF-GSRSEGASFGAPMAKDAKL 137
Db      748 SSSLSRAAGRLQPLSLR-PSAARPCRRRLRRRSRLIFAGAASNSGNNHRRRLVLAGKVAL 806
Qy      138 RALLWTPPTPTGSHSPRPRE-APLR---AHPAGS---KTEPPADDSQKLSMG--G 187
Db      807 PPLYCLQPLPRRSRGARLRLRRRRVALHPGLPSVLPERRRP-----RRLSLGLVG 860
Qy      188 LHSRPLKRLSHSLTHANVPTSGHPATSAFHTNGPDLRPESTSGVTFRSPVTSRRASV 247
Db      861 THSGSPRR-----ASAPADVAVTSIGIA-----AGKV 888
Qy      248 SISV---PSTPRRGATOKP 264
Db      889 VLGIASPPSKRLGSPSSSP 908

RESULT 14
US-10-437-963-200034

```

```

; Sequence 200034, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
US-10-437-963-200034

Query Match          9.3%; Score 133; DB 16; Length 380;
Best Local Similarity 24.2%; Pred. No. 0.067;
Matches 62; Conservative 16; Mismatches 114; Indels 64; Gaps 9;

Qy      41 SPAGTRPTPDPPPEWEKANRTG-----VGKEASKALGAKSCETTPSRG 87
Db      86 SAAAAAPPAPAPPPRRRRRCRGNPPPLPRSPSSAASAPSRASPPAPATPCPAP 145
Qy      88 STPTLTPRKKKKYRPISTPSYCDSESLFSGRSBGASFGAPR---MAKDDAKLALMT 143
Db      146 RCGSATP--PAPWRPXPSPSPSPAPAPACGCAARTPPAYRRLSVITPPPRHHAAVL 203
Qy      144 PPPTPR---GSHSPRREAPLRAIH-----PAGSKTEPPGPAADSQKLSMGSLHSRP 193
Db      204 PPAPPRRNRTTRTPRPPRRHTLRIPPPPLPAPSPAPPG----- 246
Qy      194 LKRGLSHLTHLNVPTSGHPATSAFHTNGP--QDLRPSYSGVTFRSPVTSRRASVSI 251
Db      247 -----RHPPPPSP-HPPXGPRPRPPAGNHLHSPPPPNPNPLPPSR-----GP 290
Qy      252 PSTPRRGATOKPKP 267
Db      291 PPSPSRSPATDLPAP 306

RESULT 15
US-10-437-963-111701
; Sequence 111701, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
US-10-437-963-111701

```

; SEQ ID NO 111701  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(514)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_15655C.1.pap  
; US-10-437-963-111701

Query Match 9.3%; Score 133; DB 16; Length 514;  
Best Local Similarity 26.6%; Pred. No. 0.094;  
Matches 64; Conservative 20; Mismatches 117; Indels 40; Gaps 9;

OY 42 PAGTRPPPPDPPWVEKANRTRGVQKEAS-----KALGAKGSCETTPSRGSPPTLTP 94  
DB 239 .PARTLPAPPXADPPGRGXPRRRATOTRAAPPACTRRSPAGCGCWSPPRPPGGSPTPAP 298  
OY 95 RKKNKYRPISTPSYCDESLFGSRSEGASFGAPRMAKGDALRLALMTPTPTPRGSHSP 154  
DB 299 -----PPPPPPRGAPRAAPASAAAPASAAAGRAPRAPRPPRPPPPPSARASSSP 352  
OY 155 RPREAPLRA---IHPAGSPKTEPGPAADSOQLSMGGLHSRPLKRGGLSHSLTHLNPST 210  
DB 353 RHSATPAPAHXPRRPSRGRPPPTRAAPS-----PMPPTRTGAAP-----RGPST 398  
OY 211 GHPATSAPHING-PODLR---PSTSGVTFRSPLVTSRARSVSISVSTPRRGATOKPKP 266  
DB 399 PAPSSSWPPAGTPAALRWADAPASAPGTACS--AAGRAAG---SSPPRPQHRRHHHPSPP 453  
OY 267 P 267  
DB 454 P 454

Search completed: March 28, 2005, 09:05:30  
Job time : 391 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 08:00:05 ; Search time 24 Seconds  
(without alignments)  
1078.430 Million cell updates/sec

Title: US-10-031-589-4

Sequence: 1 MKTVELAVSGMQTLGLQHR.....SVSPTRRGATQKPKPWWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*

1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144.5	10.1	539	2 T28770	hypothetical prote
2	136	9.5	2715	2 T13049	eyelid - fruit fly
3	132.5	9.2	580	2 T43481	probable mucin DKF
4	130.5	9.1	705	2 A35363	synapsin I splice
5	128	8.9	403	2 S52796	prl2 protein - hu
6	127.5	8.9	1870	2 S37671	MHC class III hist
7	127.5	8.9	1872	2 S36152	MHC class III hist
8	127	8.8	302	2 T32711	hypothetical prote
9	126.5	8.8	1199	2 A40670	nuclear envelope p
10	126	8.8	706	2 E30411	US4 protein - huma
11	125.5	8.7	699	2 C43674	probable, serine/th
12	124	8.6	576	2 T36728	hypothetical prote
13	123.5	8.6	724	2 T47149	glycoprotein 350/2
14	123	8.6	886	2 S29605	membrane antigen 9
15	122.5	8.5	907	1 QOBE21	hypothetical prote
16	122	8.5	1777	2 T34369	hypothetical prote
17	122	8.4	721	2 E70766	high molecular mas
18	121	8.4	1151	2 T18535	MHC class III hist
19	120.5	8.4	2142	2 B50398	NIMA-like protein
20	120	8.4	779	2 A57177	unconventional myo
21	120	8.4	3530	2 A59266	immediate-early pr
22	119	8.3	1446	1 A45344	En/Spm-like transp
23	118.5	8.2	1077	2 T52305	unconventional myo
24	116.5	8.1	306	2 T52305	hypothetical prote
25	116.5	8.1	3511	2 A59295	hypothetical prote
26	116	8.1	635	2 T75477	hypothetical prote
27	115.5	8.0	846	2 T21700	hypothetical prote
28	115.5	8.0	883	2 A96662	hypothetical prote
29	115.5	8.0	1285	2 T14171	ataxin-2 - mouse

30	115.5	8.0	1456	2 T01397	LTR gag/pol polypr
31	115.5	8.0	2176	2 T13806	toncan gene protei
32	115.5	8.0	2212	2 A41098	calcium channel pr
33	115	8.0	380	2 D70516	hypothetical prote
34	115	8.0	704	2 A30411	synapsin Ia - rat
35	114.5	8.0	308	2 S7938	EBNA-LP protein -
36	114.5	8.0	1211	2 T42230	AP4 protein - mous
37	114.5	8.0	2187	2 T30826	nascent polypeptid
38	114	7.9	2346	2 T13829	Trp homolog - frui
39	113.5	7.9	346	2 T46916	hypothetical prote
40	113	7.9	1952	2 T48814	hypothetical prote
41	112.5	7.8	2282	2 T42717	DNA-binding protei
42	112	7.8	214	2 T10737	extensin-like cell
43	112	7.8	214	2 T09854	proline-rich cell
44	112	7.8	620	2 S06733	hydroxyproline-ric
45	112	7.8	3149	1 QOBE8	Bpuf1 protein - hu

## ALIGNMENTS

RESULT 1  
T28770  
hypothetical protein W03D2.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28770  
R:Kohlring, T.; Wohlmann, P.  
A:Description: The sequence of C. elegans cosmid W03D2.  
A:Reference number: Z20519  
A:Accession: T28770  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-539 <ROH>  
A:Cross-references: UNIPROT:Q8M0G9; EMBL:AF000298; PIDN:AAC48255.1; GSPDB:GN00022; CESP:  
A:Experimental source: strain Bristol N2; clone W03D2  
C:Genetics:  
A:Gene: CESP:W03D2.1  
A:Map position: 4  
A:Insertions: 40/3; 88/3; 115/3; 146/3; 173/3  
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 10.1%; Score 144.5; DB 2; Length 539;  
Best Local Similarity 24.1%; Pred. No. 0.025;  
Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;

QY	13	QTLGLQHRGCGYVVKARTSYVDLFGSPAGTRPTPDPPMWEKANRRGVGEASK	72
DB	234	QGRGQOQOTONGFRGKRO-----APPAGSPPPP--P-----KGGP	268
QY	73	ALGAKGSCETPPSRGSPPTLTPRKKNKYRPISTHTPSYCDLSFGSRSGAGFGAPRMAGK	132
DB	269	PLAGSGPPPPPAAGSP--PPTGSPPPPTGSP-----PPPPAGG	308
QY	133	DAALRALMLTPPTPPGSS-----HSPPR-----REAPLRATHPA	167
DB	309	SPPPRAGSPPPPPPPGSPPTGSLPPPOAGSGPPPAAGTGPDPPPPRKQRAPEPSPTG	368
QY	168	GPSKTEP-----GPAADGSKLS-----MG-----LHSSRLKXGL	198
DB	369	SPPTGSPPTGPPPGGPKSSSESSSESRGCGGPGGPGGPKKSSSESSSESRKEPPGP	428
QY	199	SHSLTHLVNPSGTGHPATSAPTNGPQDLRPSTGVTFRSPLVTSRARVSISVSPSTPRG	258
DB	429	RRS-----PPTGSPPTGSPPTGPPGSPPTG-----SPTGLPSRQKQ	468
QY	259	GATOKP--KPP	267
DB	469	APEDRPTGSP	479

RESULT 2

T13049  
 evelid - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
 C:Accession: T13049  
 R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.  
 submitted to the EMBL Data Library, March 1998  
 A:Reference number: Z17592  
 A:Accession: T13049  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2715 <TRE>  
 A:Cross-references: UNIPROT:Q8IN94; EMBL:AF053091; NID:G2981220; PID:G2981221; PIDN:ACC  
 C:Genetics:  
 A:Gene: eld  
 A:Cross-references: FlyBase:FBgn0003013  
 C:Function:  
 A:Description: could act as a transcription factor antagonistic to the wg pathway  
 C:Keywords: DNA binding

Query Match	9.5%	Score 136	DB 2	Length 2715
Best Local	Similarity 22.5%	Pred. No. 0.47	Mismatches 119	Indels 64
Match 62	Conservative 30			Gaps 11
Qy	23	GGYRV-----KARTSYVDLTFGSDAGRPRPPDF-----DPWVEKANRRTRGVG	67	
		::		
Db	632	GGYKKGKQPGQSGAQQYPPQQPQQYPPGNYPPRPQYPPGAYATGPPPPPSQA----	GAG 687	
Qy	68	KEASALAKKSGCETTPSRGSGTPTLPRKKKXRPISHTPSYCDSELFGSRSGAAGAP	127	
		::		
Db	688	GANSPPSGAQAQ-GYFGRG-----MHNHTGGQPPYQWPPSPQQTVPFGAAGGAAVGNH	740	
Qy	128	EMAKDAKALRALLTWTP-----PTPRGSHSPREARPLRAINPDAGSKTEPPGADS	180	
		::		
Db	741	VQSK-----TPPPVVGSGPPPGSGSPRLYKHLQHKGGYGGSGPTPPQG	790	
Qy	181	OKLSMG--GLHSRPL-----KRLGSLTLHLWVSPF-----GHPATAPHT	220	
		::		
Db	791	QCYGNGPPTGMHMGKPMGPRPHHMGPRHPINMGPRPTTTPPQSQMLGCGQPGGAGASGPRS	850	
Qy	221	NGPQDLRPTSGVTRFSPLVTSRANVSISVSTP	255	
		::		
Db	851	GGPEHI-SQDNGIISSSGPTGAAGMAHVAIVSVTTGP	884	

### RESULT 3

probable mucin DKFZp434C196.1 - human (fragment)  
 N/Alternate names: protein DKFZp434B0635.1  
 C/Species: Homo sapiens (man)  
 C/Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text\_change 09-Jul-2004  
 C/Accession: T43481, T34549, T17264  
 R/Kocher, K.; Beyer, A.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, December 1999  
 A/Accession: T43481  
 A/Molecule type: mRNA  
 A/Residues: 1-580 <AA>  
 A/Cross-references: UNIPROT:Q9UF83; EMBL:AL13561; NID:g65959133; PIDN:CAB63715.1, PID:g65959133  
 A/Experimental source: adult testis; clone DKFZp434C196  
 R/Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, October 1999  
 A/Reference number: Z21540  
 A/Accession: T34549  
 A/Molecule type: mRNA  
 A/Residues: 262-580 <POU>  
 A/Cross-references: EMBL:AL122069; NID:g6102864; PIDN:CAB59245.2; PID:g7018420  
 A/Experimental source: adult testis; clone DKFZp434B0635  
 R/Poustka, A.; Klein, M.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, September 1999  
 A/Reference number: Z18723  
 A/Accession: T17264  
 A/Molecule type: mRNA

A<sub>1</sub>Residues: 262-580 <P0U2>  
A<sub>1</sub>Cross-references: EMBL:AL117481; NID:g5911958; PIND:CAB55054.1; PID:g5911959  
A<sub>1</sub>Experimental source: adult testis; clone DKFZp34B061  
C<sub>1</sub>Genetics:  
A<sub>1</sub>Note: DKFZp34C19.6.1; DKFZp43B0635.1; DKFZp43B061.1

Query Match	9.2%	Score 132.5	DB 2	Length 580
Best Local Similarity	26.3%	Pred. No. 0.16		
Matches 77; Conservative	30; Mismatches 125;			Indels 61; Gaps 17

```

Oy      3 TPVELAVSGMOTLGLOHRCRGYRVKARTSYV-----DETLEFGSPAGTRPT--PPDEDPW 56
          || ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      46 TPFRASLTKEHSTAL-----LRTLPRASLMRTPTRASLMRTPPRASPTPKPPRASPT 90

```

```

Oy      57 VEKANTRGVGEKASKALGAKSCETTPSRGSGPTLPKKNKYRPI SHPPSYCBESLFG 11
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      99 PSRASPTRLRLRASPMGSPHRRASPMRTTPRAS -PTGTPTASP----TGTPS--SASPTG 15

```

```

QY      117 SRSEGSGFGADPRMA----KGDANKL-----RALIWTPTPTPTRGSHSPRPEEAPLAIHPA 16
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      152 TPTRASPTGTTPRAWATRSPTASLTTRTPSRASLTWMPRASPTKPT-PRSP-RNSHRA 200

```

```
Qy      168 GPSKTEP--GPAADSOGLSMGGHSSNP LKRLG--SHSLTHLWNPSTGHATSAPHTNGP 22
       |||::|||::|||::|||::|||
Db      210 SPTRTTPRASPTRRPPRAS----PTRTTTPRESLTSHSASPTRNPPRASPTRRPP----- 26
```

```

Qy      224 QDLRPSTSGVTR-SPLVTSRARSVISVPSTRRGATQPK-----ppw 268
          | : | | | | | | | | | | | | | | | | | | | |
Db      261 --KASPTGSPPRASPMTPPRAS-----PRTPPRASPTTPSRASLRTPSW 304

```

## RESULT 4

synapsin I splice form a - human  
C1Species: Homo sapiens (man)  
C1Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 09-Jul-2004  
C1Accession: A35363; B35363; A35805  
R1Suedhof, T.C.  
J. Biol. Chem. 265, 7849-7852, 1990  
J1Title: The structure of the human synapsin I gene and protein.

A/Accession: A35363  
A/Molecule type: DNA  
A/Residues: 1-705 <SUE>

A:Accession: B35363  
A:Status: preliminary  
A:Molecule type: DNA

A: Cross-references: GB: M58378; GB: J05431  
R: Sauerwald, A., Hoesecke, C., Oeschwald, R., Kilian, M.W.  
J. Biol. Chem. 265, 14932-14937, 1990

A:Reference number: A35805; MUID:90368667; PMID:2118519  
A:Accession: A35805  
A:Molecule type: DNA

ACross-references: GB:M5301; NID:g338655; PIDN:AAA60608.1; PID:g553654; GB:J05630  
 CGeneicgb:  
 AIGene: GDB:SYN1

A/Map position: Xp11.23-Xp11.23  
C/Keywords: actin binding; alternative splicing; phosphoprotein

Best Local Similarity	23.5%	Pred. No. 0.27;
Matches	61;	Conservative
	24;	Mismatches
	108;	Indels
	67;	Gaps
	9	

Db 454 GPRARPR-PRGGPRQRGRGRRLQRRPRRGGQNHSLGRPRG-SPLRQLRSP 511111

Db 512 TSAPO-----OPASQAAAPPTGCGRQSRPVAGPGAPPAAPASPSPORQAGPQATRO 566  
Qy 150 -----GSHSPRRERAPLRAIHAPGPKTERGPAADSOQLSMGSLHSRLXGLSHSLTHL 205  
Db 567 TSVSGPAPPPKASGAPPGGQORQGPQKPPGAPGAPRQASQAG----- 608  
Qy 206 NVPSGTGHATAPHTNGP-----ODLRPSTSGVTFRSP---LVTSRARSV 247  
Db 609 PVPRTPPTTQQPRPSGGRGAPAPKPKQALQKPSQDVPPAPPAAGGPHPOLNLSQSILTN 668  
Qy 248 SISVSTPRRGATQKPP 267  
Db 669 AFNLP-----EPAP 678

RESULT 5  
S52796  
Prp2 protein - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-May-1995 #sequence\_revision 10-Nov-1995 #text\_change 10-Sep-1997  
C:Accession: S52796  
R:Ruhmann, A.; Kreideweis, S.; Nordheim, A.  
submitted to the EMBL Data Library, March 1995  
A:Reference number: S52796  
A:Accession: S52796  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-403 <RUI>  
A:Cross-references: EMBL:X86019; NID:9762950; PID:9762951

Query Match 8.9%; Score 128; DB 2; Length 403;  
Best Local Similarity 25.3%; Pred. No. 0.22; Indels 52; Gaps 11;  
Matches 62; Conservative 23; Mismatches 108;  
Qy 42 PAGR-PTPPDPPEWEKANRTGKASALGAKSCETTPSRGTP--LTPRKN 98  
Db 116 PGGRQSPGPTPPPPGNRGTALGGGSIROSPLSSSPFNRPPLPTPRALDDKPP 175  
Qy 99 KYRISTHTSYCDLSLFGSRSEGASFGAPMAKGDAAKLLMTPTTPPGSHSPRR- 157  
Db 176 PPPVGNRPSTHRAV-----PPPPQNNKPPVPST 206  
Qy 158 ---BAPLRA-IHPAGSKTERGPAADSOQLSMGSLHSRLXGLSHSLTHLVNSTGHP 213  
Db 207 PRBAPHRHLPAPPPSPGPPPLPPS---SSGADETRLPQKNISSLSTPPSPERS 263  
Qy 214 A---TSAHTNGPODLR--PSTSGVTFRSPVTSRARSVSVSTPSP---RGATQK-- 263  
Db 264 GPLPPVPSERPVPNDPPRSG-PLPPPPVSRNGSTSLALPATLPQSPRSGVDSPRS 322  
Qy 264 -PKPP 267  
Db 323 GPRPP 327

RESULT 6  
S37671  
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human  
C:Species: Homo sapiens (man)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 15-Sep-2000  
C:Accession: S37671  
R:Bouguieret, L.  
submitted to the EMBL Data Library, August 1992  
A:Reference number: S37671  
A:Accession: S37671  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1870 <BOU>  
A:Cross-references: EMBL:Z15025; NID:929374; PID:929375  
C:Genetics:  
A:Map position: 6p21.3  
A:Insertions: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65

C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.9%; Score 127.5; DB 2; Length 1870;  
Best Local Similarity 25.2%; Pred. No. 1.1; Indels 73; Gaps 12;  
Matches 68; Conservative 17; Mismatches 112;  
Qy 33 YVDETLF-GSPAGTRPTPPDPDPV-----EKANTRGVGKE-----ASKALGAKS 79  
Db 728 YVDETLFGR-----PLDTPPGVHPSGLVPRERSRSLSEPPRHAAPMLREKGT 781  
Qy 80 CETTPSR---GSTPTLTPRKKNKTRPI-SHTPSYCDLSLFGSRSEGASFGAPMAKGDAA 135  
Db 782 PPVDPKLAWGDVFTATPAEP---RPLTSPRLQAADDDKGRSE----- 823  
Qy 136 KLRALLMTPTTPPGSHSPRRERAPLRAIHAPGSKTERGPAADSOQLSMG----- 187  
Db 824 -----TPPV-----PP-----PYLASYGFPENGTPGPPIRFPLEEGPRELPMP 867  
Qy 188 -----LHSPRLKRLSHSLTHLVNSTGHPATSAHTNGPODLRPSGTGTFRSPV 240  
Db 868 GSDVAKIQTPPPKPEKPEKETAQLTGPRAGKLPASRSGAGPPPRRSEKTRMGPRP 927  
Qy 241 TSRAASVSIYP-STPRRGATQKPPWK 269  
Db 928 GSSRRGIPPEERGAAPRRAGPIKKPPPTK 957

## RESULT 7

S36152  
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Jun-1995 #sequence\_revision 17-Nov-1995 #text\_change 15-Sep-2000  
C:Accession: S36152  
R:Riris, F.U.M.; Bouguieret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Jurka  
Nature Genet. 3, 137-145, 1993  
A:Title: Dense Alu clustering and a potential new member of the Nfkapab family within a  
A:Reference number: S36152; MUID:93272029; PMID:8499947  
A:Accession: S36152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1872 <RI>  
A:Cross-references: EMBL:Z15025  
A>Note: in the authors' translation residues 32-34 are shown after residue 4 and, conseq  
A:Insertions: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.9%; Score 127.5; DB 2; Length 1872;  
Best Local Similarity 25.2%; Pred. No. 1.1; Indels 73; Gaps 12;  
Matches 68; Conservative 17; Mismatches 112;  
Qy 33 YVDETLF-GSPAGTRPTPPDPDPV-----EKANTRGVGKE-----ASKALGAKS 79  
Db 729 YVDETLFGR-----PLDTPPGVHPSGLVPRERSRSLSEPPRHAAPMLREKGT 782  
Qy 80 CETTPSR---GSTPTLTPRKKNKTRPI-SHTPSYCDLSLFGSRSEGASFGAPMAKGDAA 135  
Db 783 PPVDPKLAWGDVFTATPAEP---RPLTSPRLQAADDDKGRSE----- 824  
Qy 136 KLRALLMTPTTPPGSHSPRRERAPLRAIHAPGSKTERGPAADSOQLSMG----- 187  
Db 825 -----TPPV-----PP-----PYLASYGFPENGTPGPPIRFPLEEGPRELPMP 868  
Qy 188 -----LHSPRLKRLSHSLTHLVNSTGHPATSAHTNGPODLRPSGTGTFRSPV 240  
Db 869 GSDVAKIQTPPPKPEKPEKETAQLTGPRAGKLPASRSGAGPPPRRSEKTRMGPRP 928  
Qy 241 TSRAASVSIYP-STPRRGATQKPPWK 269  
Db 929 GSSRRGIPPEERGAAPRRAGPIKKPPPTK 958

## RESULT 8

T32711  
 Hypothetical protein T22D1.2 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T32711  
 R/Gelstein, C.; Bradshaw, H.; Hawkins, M.  
 submitted to the EMBL Data Library, December 1997  
 A/Description: The sequence of *C. elegans* cosmid T22D1.  
 A/Reference number: Z21211  
 A/Accession: T32711  
 A/Status: preliminary; translated from GB/EMBL/DDBJ  
 A/Molecule type: DNA  
 A/Residues: 1-302 <GB>  
 A/Cross-references: UNIPROT:Q96ZM1; EMBL:AF039052; PIRIN:AAB94271.1; GSPDB:GN00022; CESP:  
 A/Experimental source: strain Bristol N2; clone T22D1  
 C/Genetics:  
 A/Gene: CESP:T22D1.2  
 A/Map position: 4  
 A/Introns: 89/1  
 C/Superfamily: proline-rich protein

Query Match 8.8%; Score 127; DB 2; Length 302;  
 Best Local Similarity 24.6%; Pred. No. 0.19;

Matches 65; Conservative 15; Mismatches 98; Indels 86; Gaps 11;

QY 44 GTRPTPDPPPP--WZKAKRTR-----GVGKSKALGAKG--SCETTPRGSTPTL 92  
 DB 72 GTRPPPTGPPGPDLSAEAGNARRPPPPKGTGTPPPPTGEPDLSGEGNARRPP-- 129  
 QY 93 TPRKKNKYPRIHSHTPSYCDLSLFGSRSEGASFGAPRAKGDAAKLRALLWTPPTPRGSH 152  
 DB 130 -PPKGTSSPPPPPT-----GEPDLSGEGNASR-----RPPPPKGTG 166  
 QY 153 SPSP-----REAPLRHAPAGSKTEPPPADSQTLNMGGLHSRPLK 195  
 DB 167 SPSPPTGEPDLSGEGNARRPP--PPKGTGTPPPPTGEPDLSAEVARRP-- 220  
 QY 196 RGLSHSLTHLVNVTGHPAISAPHTNGPDLR-----PSTGVTRSPLYTSRAR 245  
 DB 221 -----PPPKGSGSPPTGTPGPDLSGEGNARRPPPPKGTGSPPPPTGEPQ 270  
 QY 246 SVSISVPTPRRGATOKPKPPMK 269  
 DB 271 DLS-----GEGNASRRPPPPK 287

## RESULT 9

A40670  
 nuclear envelope protein POW 121 - rat  
 C/Species: *Rattus norvegicus* (Norway rat)  
 C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
 C/Accession: A40670  
 R/Haliberg, E.; Wozniak, R.W.; Blobel, G.  
 J. Cell Biol. 122, 513-521, 1993  
 A/Title: An integral membrane protein of the pore membrane domain of the nuclear envelope  
 A/Reference number: A40670; MUID:93328754; PMID:8335683  
 A/Accession: A40670  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-1199 <HAL>  
 A/Cross-references: UNIPROT:P52591; GB:Z21513; NID:G396746; PIRIN:CAA79725.1; PID:G396747  
 F/803-807,845-849,956-960,1010-1014,1047-1051,1076-1080/Region: pentapeptide motif (X-F-

Query Match 8.8%; Score 126.5; DB 2; Length 1199;  
 Best Local Similarity 26.1%; Pred. No. 0.84;

Matches 67; Conservative 26; Mismatches 131; Indels 33; Gaps 11;

QY 28 KARSYVDLTF-----GSPACTRPTPDPPPWZKAKRTRGVGKSKALGAKGSCETT 83  
 DB 362 RSRRTSVSILSTCTGSISSRNA-----ITSSYSTRGV-SQLMKSGPTSSPSS 413  
 QY 84 P--SRGSTPTLTPRKKNKYPRIHSHTPSYCDLSLFGSRSEGASFGAPRAKGDAAKLRALL 141

DB 414 PASSRQTEP-RPAKTRREBPCHQSSSAPLYTDKESPEKVTDPATKQSS-----L 466  
 QY 142 WTPPPPTPRGSHSREAPL-----RAIHAGSKTEPPGAASQKLSMGGLHSRPLKRG 197  
 DB 467 WTSPPPT-GSSGQKKRIKILPSRRDQTLTPPPPLGYSITAEEDIMERRASLQWFNKY 525  
 QY 198 LSHSLTHLVNVTG--HPATSAPHTNGPDLRSTSGVTF-----RSPLVTSRAVSISVP 252  
 DB 526 LEPKTDASTPATDTSPTSPPTTLTPVVGPAASASLPAPSSNPLLSIKMKMGSSPAP 585  
 QY 253 ST--PPRGATOKPKP 267  
 DB 586 SSSEPEEATVAAPSP 602

## RESULT 10

E30411  
 synapsin Ia - bovine  
 C/Species: *Bos primigenius taurus* (cattle)  
 C/Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 01-Dec-2000  
 C/Accession: E30411; F30411; A35758  
 R/Suedhof, T.C.; Czernik, A.D.; Kao, H.T.; Takei, K.; Johnston, P.A.; Horiuchi, A.; Kana  
 Science 245, 1474-1480, 1999  
 A/Title: Synapsins: mosaic of shared and individual domains in a family of synaptic vesic  
 A/Reference number: A30411; MUID:89388265; PMID:2506642

A/Accession: E30411  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-706 <SUE>  
 A/Accession: F30411  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-664, 'KA', 664, 'PAQAP', <SUE>  
 R/Hall, F.L.; Mitchell, J.P.; Vulliam, F.R.  
 J. Biol. Chem. 265, 6944-6948, 1990  
 A/Title: Phosphorylation of synapsin I at a novel site by proline-directed protein kinase  
 A/Reference number: A35758; MUID:90216728; PMID:2108963  
 A/Accession: A35758  
 A/Molecule type: protein  
 A/Residues: 532-556 <HAL>  
 C/Keywords: actin binding; alternative splicing; phosphoprotein  
 F/551/Binding site: phosphate (Ser) (covalent) (by proline-directed kinase) #status exper

Query Match 8.8%; Score 126; DB 2; Length 706;  
 Best Local Similarity 22.9%; Pred. No. 0.52;

Matches 60; Conservative 23; Mismatches 87; Indels 92; Gaps 12;

QY 40 GSPAGTRPTPDPPPWZKAKRTRG-----VGEKSKALGAKGSCETTPSRGSTP 90  
 DB 454 GPPAQQR-PPGCGPPQPGPQRPPLQGRTPPGQGHLSGLGPPAG--SPPLRLPSP 511  
 QY 91 TTPRKKNKYPRIHSHTPSYCDLSLFGSRSEGASFGAPRAKGDAAKLRALLWTP--PPTP 148  
 DB 512 TSVPO-----QPASQ-----HTPMTQGGQSRPVPAGGPAVPAT 546  
 QY 149 RGSHPRR-----EAPLRA--IHPA-----GPSKTEPPADSOQLSM 185  
 DB 547 RPPASPPQQAQPPQATROTTSVGAHPPAASVPFGGQGRQPPQKPPAPPTQAQSQ 606  
 QY 186 GGLHSRPLKRLGSLHSLTHLVNVTGHPAISAPHTNGPDLRSTSGVTRSPLYTSRAR 245  
 DB 607 AG-----PMERTGPPTTQGRPSGPGAPGRPTKQLAQK----- 641  
 QY 246 SVSISVPTPRRGATOKPKP 267  
 DB 642 SODVPPPATMAAG-----PP 657

## RESULT 11

C43674  
 USA protein - human herpesvirus 2 (strain HG52)  
 C/Species: human herpesvirus 2



C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C/Accession: C43674  
R/McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.  
J. Gen. Virol. 68, 19-38, 1987  
A>Title: DNA sequence and genetic content of the HindIII 1 region in the short unique co  
national comparisons.  
A/Reference number: A43674; MUID:87111457; PMID:1027242  
A/Accession: C43674  
A/Molecule type: DNA  
A/Status: preliminary  
A/Residues: 1-699 <MCG>  
A/Cross-references: UNIPROT:P13390; EMBL:X04798; NID:G59900; PID:G59904  
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 8.7%; Score 125.5; DB 2; Length 699;  
Best Local Similarity 23.0%; Pred. No. 0.56;  
Matches 59; Conservative 27; Mismatches 92; Indels 79; Gaps 11;

Qy 35 DETLFGSPAGTRPPEDPPVVEKANRTGVGKEASAKALGKSGCETTPSRGTPPTLP 94  
Db 405 EETAVASPPATASV-----SSLPAAAAATPGAGHTNTSSAAKTPPTPA-----PTTPP 457  
Qy 95 RKKKKRPIPSHTPSYCDLSFGSRSESGASFGAPMAKGDAAKRLALMTTP-----PTPRG 150  
Db 458 PTSTHATPRPTT-----GPGT-----TPPGPATPGVPG 486  
Qy 151 SHSRPRAPRALIHPAGPSKTEPPADSOGLSMGSLHSSRLKRGSHS-----LTHL 205  
Db 487 ASAAPLADSPPLTASPPA---TAPGSAANVSVA---ATTATPGTGRGATRPPTDPKTHP 539  
Qy 206 NVPSGTGHPATAP-----HTNGPQDUR-----PSTGVTFRSPPLVTSRAVSIS 250  
Db 540 HGPADAPPGSPAPPEPRRGGEEREGADGEPEDDDSATGLARTENPK----- 591  
Qy 251 VPSTPRRGATQKPKP 267  
Db 592 --PPAPRGPIRPLTP 606

RESULT 12  
T36729  
probable serine/threonine-specific protein kinase (EC 2.7.1.1) - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Aug-2004  
C/Accession: T36729  
R/Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1999  
A/Reference number: Z21612  
A/Accession: T36729  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-576 <MUR>  
A/Cross-references: UNIPROT:Q9XA04; EMBL:AL079308; PID:CA45227.1; GSPDB:GN00070; SCOPED  
A/Experimental source: strain A3(2)  
C/Genetics:  
A/Gene: SCOPED:SC69\_30  
C:Superfamily: protein kinase homology  
C/Keywords: phosphotransferase; serine/threonine-specific protein kinase

Query Match 8.6%; Score 124; DB 2; Length 576;  
Best Local Similarity 26.8%; Pred. No. 0.57;  
Matches 74; Conservative 20; Mismatches 90; Indels 92; Gaps 19;

Qy 41 SPAGTRPPDPDPVVEKANRTGVGKEASAKALGKSGCETTPSRGTPPTLP 98  
Db 319 SPAPTRPPDPVVEKANRTGVGKEASAKALGKSGCETTPSRGTPPTLP 368  
Qy 99 KYRPISTHPSYCDLSFGSRSESGASFGAPMAKGDAAKRLALMTTPPTPRGSHSPRRE 158  
Db 369 PGTVATATGTSAGLPASDOG-----WT-PSTPGGPTA--PPS 405  
Qy 159 APLRAIHGAPSKTEPPADSOGLSMGSLHSSR-PL-KRGLSHSLT---HLNVPS----- 209

Db 406 APSAPSAPSPCTRPAP-----HGTSEEVPLAERPGAMSETGSPHLPPOPTVT 456  
Qy 210 -TGHATTS-----APR---TNGPQDLPSPIS-----GVTPRSPLVTSR 243  
Db 457 PTDAAADDAAMAAQPPAPHPAFGGRGPPDRAPRSGHPAPHPPLTARS-LAPSP 515  
Qy 244 ARSVI-----SVPTPRRG-ATQKPKP 267  
Db 516 ARRADVPYTAATYARNRPSAPPAQHRGARRRRRPGPP 551

RESULT 13  
T47149  
hypothetical protein DKFZps47p103.1 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 16-Aug-2004  
C/Accession: T47149; E56695  
R/Bloecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A/Reference number: Z24376  
A/Accession: T47149  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-724 <AAA>  
A/Cross-references: UNIPROT:Q04727; EMBL:AL162059  
A/Experimental source: fetal brain; clone DKFZps47p103  
R/Stifant, S.; Blaumüller, C.M.; Redhead, N.J.; Hill, R.E.; Artavanis-Tsakonas, S.  
Nature Gene. 2, 119-127, 1992  
A>Title: Human homologs of a Drosophila enhancer of split gene product define a novel fa  
A/Reference number: A56695; MUID:93265135; PMID:1303260  
A/Accession: E56695  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 277-418, 'P', 420-459, 'A', 461-724 <STI>  
A/Cross-references: GB:M9439; NID:G307515; PID:AAA61195.1; PID:G307516  
C/Genetics:  
A/Gene: GDB:TLE4; ESG  
A/Cross-references: GDB:228050  
A/Note: DKFZps47p103.1  
C:Superfamily: WD repeat homology  
C/Keywords: nucleus  
F:480-513/Domain: WD repeat homology <WD1>  
F:566-599/Domain: WD repeat homology <WD2>  
F:648-681/Domain: WD repeat homology <WD3>  
F:689-722/Domain: WD repeat homology <WD4>

Query Match 8.6%; Score 123.5; DB 2; Length 724;  
Best Local Similarity 23.2%; Pred. No. 0.78;  
Matches 67; Conservative 38; Mismatches 113; Indels 71; Gaps 13;

Qy 3 TPVEL-NAVSMQTLGLQHRCRGYRVKARTSYVDETLFGSPAGTRPPDPVVEKAN 61  
Db 71 TMAELNMIIGQOOLQAQHLSHG-----HGLVPLTPPHPSGLQPAIPIG 115  
Qy 62 RTRGVGKEASAKALGKSGC-----ETTPSRGSPPTLPKKNKYRPIIS 104  
Db 116 SSAGL-LALSALLAQSHSLPIKDEKKNHNDHQDRSISVSPASFGAEGHR--- 171  
Qy 105 HTPSYCDLSFGSRSESGASFGAPMAKGDAAKRLALMT-----PPTPRGS--HSPR----- 155  
Db 172 NSADYSSSK-KQKTEKEIARAYDSDEKSDNLVVDVSNEDSSPPGSPAHSPREGL 230  
Qy 156 -----PREAPLAIHPAGPSKTEPPADSOGLSMGSLHSSRLKRLSHSLHLNVPS 209  
Db 231 DKTRILKKDAP---ISPASIASSSSTPSKSKELSLN-----EKSTTPVSKSN 275  
Qy 210 TGHATSA--PTNGPQDLPSTSGVTPFRSPPLVTSRAVSISVP-STP 255  
Db 276 TTPRTIDAPTPGNSSTGLRVPCKPPGVDPDLASSLRTPAVPCPYTP 324

RESULT 14  
S29605

glycoprotein 350/220 - human herpesvirus 4  
 C:Species: human herpesvirus 4, Epstein-Barr virus  
 C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Jul-2004  
 C:Accession: S29605  
 R:Klein, R.; Mueller-Lantzsch, N.  
 A:Submitted to the EMBL Data Library, October 1992  
 A:Description: Sequences of the membrane proteins gp 350/220 and p140 of Epstein-Barr vi  
 A:Reference number: S29605  
 A:Accession: S29605  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-886 <KLE>  
 A:Cross-references: UNIPROT:Q07284; EMBL:X67776; NID:959163; PIDN:CAA47986.1; PID:959164  
 C:Superfamily: Epstein-Barr virus membrane antigen gp350  
 C:Keywords: glycoprotein

Query Match 8.6%; Score 123; DB 2; Length 886;  
 Best Local Similarity 23.9%; Pred. No. 1; Mismatches 120; Indels 102; Gaps 15;  
 Matches 79; Conservative 30;

```

QY 3 TPVELAVSGMOTLGLQHRCRGGRVYKARTSYVDLTFSGSPAGTRPT-----PDPPPW 56
DB 529 TPTPNATS--PTLG-----KTSPTSAVTTPTPNATSPPTAVTTPTPNATIPR 573
QY 57 VEKANRTG-----VKEASKA-----LGAKSGCE--TTPSGSTPTLPRK 96
DB 574 LGKTSPTSAVTTPTPNATSPPTVGETSPQANNTNHTLGSTSPVVTSPKANATSAVTTGQ 633
QY 97 KN-----KRPISHTPSYCDSELFGSRSEGASFGAPRNAGK-----DAAKLRAL 140
DB 634 HNTSSSTSSMSLRPSSISFTLSPTSDNSTSHMPLTSAHPGGENITQVTPASSTTH 693
QY 141 LMTPEPTPT-----GSHSPRPREA-----PLRAHPAGPS-----KTE 173
DB 694 VSTSSPARPPTGTSQASGPGNSSTKGEVNVTKGTPPKNATSPQAPSGGKTAVPTVTS 753
QY 174 PGPAADSOGLSMGILHSRPLKRGLSHSLTHLNPST---GHPATSAHPGQDLRPST 230
DB 754 TCGKANS---TTGGKHTT-----GHGARTSTETPTDYGDSSTTPRTRYNATVTLPPST 803
QY 231 SGV-----TFRSPLVTSRARSVSISVPSTPR 256
DB 804 SKLRPRWTFSTSPVTTAQTAVPVPTSQPR 834

```

## RESULT 15

OQBE21

membrane antigen gp350 - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4, Epstein-Barr virus

C&gt;Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 09-Jul-2004

C:Accession: A43042; S33008; S33009; A03762

R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A&gt;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A:Reference number: A93065; MUID:85035713; PMID:6092825

A:Accession: A43042

A:Molecule type: DNA

A:Residues: 1-907 &lt;BAN&gt;

A:Cross-references: UNIPROT:P03200; GB:V01555; GB:J02070; GB:K01729; GB:K01730; GB:V0155

R:Farrell, P.J.

submitted to the EMBL Data Library, March 1988

A:Reference number: S32973

A:Accession: S33008

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 &lt;PAR&gt;

A:Cross-references: EMBL:V01555; NID:959074; PIDN:CAA24855.1; PID:91334869

A:Accession: S33009

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 &lt;PA2&gt;

A:Cross-references: EMBL:V01555; NID:959074; PIDN:CAA24855.1; PID:91334869

R:Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984  
 A>Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
 A:Reference number: A03794; MUID:84270667; PMID:6087149  
 A:Contents: annotation; protein coding region  
 C:Superfamily: Epstein-Barr virus membrane antigen gp350

Query Match 8.5%; Score 122.5; DB 1; Length 907;  
 Best Local Similarity 22.8%; Pred. No. 1.1; Mismatches 128; Indels 109; Gaps 13;  
 Matches 77; Conservative 24;

```

QY 3 TPVELAVSGMOTLGLQHRCRGGRVYKARTSYVDLTFSGSPAGTRPT-----P 49
DB 543 TPTPNATS--PTLG-----KTSPTSAVTTPTPNATSPPTAVTTPTPNATIPR 587
QY 50 PDPPPWVEKANRTGVEKASKALG-----AKSGCETTPERGSTPTLPRKKNKY 100
DB 588 PNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQANNTNHTLGSTSTPVVTSPKANAT 647
QY 101 RPI-----SHTPSYCDSELFGSRSEGASFGAPRNAGKDAAKLRALLMTTP 145
DB 648 SAVTTGQHNTSSSTSSMSLRPSSNPFTLSPTSDNSTSHMPLTSAHPGGENITQVTP 707
QY 146 -----PTPR-----GSHSPRPREA-----PLRAHPAGPS----- 170
DB 708 ASISTHNVSTSSPARPPTGTSQASGPGNSSTKGEVNVTKGTPPONATSPQAPSGGKT 767
QY 171 -----KTEPGPADSOGLSMGILHSRPLKRGLSHSLTHLNPST---GHPATSAHPNGP 223
DB 768 AVPTVTSSTGKANS---TTGGKHTT-----GHGARTSTETPTDYGDSSTTPRTRYNAT 817
QY 224 ODLPSTSGV-----TFRSPLVTSRARSVSISVPSTPR 256
DB 818 TYLPSTSSKLRPRWTFSTSPVTTAQTAVPVPTSQPR 855

```

Search completed: March 28, 2005, 08:53:43

Job time: 28 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2005, 05:59:19 ; Search time 80 Seconds  
(without alignments)  
1721.866 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437

Sequence: 1 MKTEVELAVSGMKQTGLGQHR.....SVSPRRRGATKPKPKPWK 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1437	100.0	269	2	Q96K30
2	867.5	60.4	253	2	Q96K30
3	863	60.1	177	2	Q96K25
4	150	10.4	647	2	Q8S148
5	144.5	10.1	518	2	Q8MG08
6	144.5	10.1	524	2	Q02123
7	144.5	10.1	539	2	Q8MG09
8	144	10.0	1067	2	Q6NV05
9	142.5	9.9	964	2	Q26963
10	140	9.7	415	1	STYL CANPA
11	138.5	9.6	659	2	Q6C708
12	136.5	9.5	2185	2	Q75M06
13	136.5	9.5	4911	1	MIL3 HUMAN
14	136	9.5	769	2	Q9L071
15	136	9.5	2556	2	Q7A5E8
16	136	9.5	2716	1	OSA DROME
17	134.5	9.4	503	1	WAIP HUMAN
18	134	9.3	463	2	Q6AVC5
19	134	9.3	1131	2	Q9VLM2
20	133.5	9.3	1303	2	Q78Y42
21	133	9.3	862	1	MCDL RAT
22	133	9.3	1240	2	Q9DM07
23	132.5	9.2	488	2	Q87W07
24	132.5	9.2	506	2	Q69125
25	132.5	9.2	580	2	Q9UFB3
26	132.5	9.2	3288	2	Q775D9
27	132	9.2	356	2	Q9Q0B7
28	132	9.2	356	2	Q9Q0B8
29	132	9.2	1374	2	Q6LA41
30	131	9.1	3326	2	Q77591
31	130.5	9.1	450	2	Q6MYW6

32	130.5	9.1	510	2	Q6WZU9	Q6mz9 homo sapien
33	130.5	9.1	705	1	STYL HUMAN	P17600 homo sapien
34	130.5	9.1	917	2	Q873D8	Q873d8 neopspora
35	130	9.0	546	2	Q7P212	Q7P212 anopheles
36	130	9.0	1725	2	Q621F2	Q621f2 burkholderi
37	129.5	9.0	506	2	Q8AZK7	Q8azk7 human herpe
38	129.5	9.0	863	1	MIL3 HUMAN	Q8m3f8 homo sapien
39	129.5	9.0	890	2	Q8UZE1	Q8uzel cercopithec
40	129.5	9.0	1014	2	Q9PUJ8	Q9flu8 oryza sativ
41	129.5	9.0	1196	2	Q6NRV6	Q6nrV6 xenopus lae
42	129	9.0	356	2	Q9Q0B5	Q9q0b5 human herpe
43	128.5	8.9	870	1	MIL3 MOUSE	Q8b9t6 mus musculu
44	128	8.9	356	2	Q9Q0B2	Q9q0b2 human herpe
45	128	8.9	356	2	Q9Q0B3	Q9q0b3 human herpe

#### ALIGNMENTS

RESULT 1  
Q96K30 PRELIMINARY; PRT; 269 AA.  
ID Q96K30;  
AC Q96K30;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Hypothetical protein FLJ14627 (Hypothetical protein PSEC0043).  
GN Name=FLJ14627;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obaeashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,  
RA Niimiya K., Ishibashi T., Yamashita H., Murakawa M., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Houta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
RA Togojya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Museshino K., Yuki H., Oshima A., Sawaki N., Motekawa S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Tagashira K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaemura T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RI Nat. Genet. 36:40-45(2004).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strusberg R. L., Feingold E. A., Grouse L. H., Derge J. G.,  
RA Klausner R. D., Collins F. S., Wagner L., Shenmen C. M., Schuler G. D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,  
 RA Kzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Oca T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayaishi K., Ishii S.,  
 RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
 RA Nagahara K., Sugano S., Isogai T.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK027733; BAB55328.1; -  
 DR EMBL; BC020922; AAH2092.1; -  
 DR EMBL; AK075358; BAC1568.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 269 AA; 28619 MW; 3F053E1454F60773 CRC64;  
 Query Match 100.0%; Score 1437; DB 2; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 2,8e-81;  
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKTPVELAVSGMOTLGLQHRCRGGRYKARTSYVDLTFSPSPGTRTPPPDPPEWYKA 60  
 DB 1 MKTPVELAVSGMOTLGLQHRCRGGRYKARTSYVDLTFSPSPGTRTPPPDPPEWYKA 60  
 QY 61 NRTRGVKEKSKALGANGSCETTPSRGSPPTLTPPKKKRTPISHTPSYCDSEILFGSRSE 120  
 DB 61 NRTRGVKEKSKALGANGSCETTPSRGSPPTLTPPKKKRTPISHTPSYCDSEILFGSRSE 120  
 QY 121 GASFGAPRAKGDAAKLRALLMTPTTPPSRSHSPRPREAPLRAIHAPGSKTEPGPAADS 180  
 DB 121 GASFGAPRAKGDAAKLRALLMTPTTPPSRSHSPRPREAPLRAIHAPGSKTEPGPAADS 180  
 QY 181 OKLSMGGLHSSRLKGLSHSLTHLNPSTGHPATSNAPHTNGODLRPSISGVTFRSPPLY 240  
 DB 181 OKLSMGGLHSSRLKGLSHSLTHLNPSTGHPATSNAPHTNGODLRPSISGVTFRSPPLY 240  
 QY 241 TSPARSVSISVPTPRRGATQKPKPFWK 269  
 DB 241 TSPARSVSISVPTPRRGATQKPKPFWK 269  
 RESULT 2  
 Q9D1H0 PRELIMINARY; PRT; 253 AA.  
 ID Q9D1H0  
 AC Q9D1H0;  
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
 DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)  
 DE Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched  
 DE library, clone=1110008J03 product:hypothetical protein, full insert  
 DE sequence (RIKEN cDNA 1110008J03).  
 GN Name=1110008J03Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=92279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:119-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA The FANTOM Consortium;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Komno H., Okazaki Y., Muramatsu M., Hayaishizaki Y.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Komno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto K., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiyama S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M.,  
 RA Onoe Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayaishizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=2288257; PubMed=12477932; DOI=10.1073/pnae.242603899;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiroka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kaubukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Komno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayaishizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;  
 RX MEDLINE=2288257; PubMed=12477932; DOI=10.1073/pnae.242603899;  
 RA Strausberg R.L., Feingold E.A., Grove L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Nitomija, Ishibashi T., Yamaashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kinata M., Watanabe M., Hiraoka S., Chiba Y., Iehida S.,  
RA Ono Y., Takisuchi S., Watanabe S., Josida M., Houta T., Kuseno J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Mueashino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro A., Tanigami A., Fujitawa T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mitsuhashi-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yamaashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs." ;  
RL Nat. Genet. 36:40-45 (2004) ;  
DR EMBL; AK027741; BAB55333.1 ; -  
SQ SEQUENCE 177 AA; 19050 MW; 834FD0C7095AED0B CRC64;

Query Match 60.1%; Score 863; DB 2; Length 177;  
Best Local Similarity 99.4%; Pred. No. 4.8e-46;  
Matches 160; Conservative % 1; Mismatches 0; Indels 0; Gaps 0

QY 1 MKTVEVLAVSGMOTLGIQHRRCGGYRYKARISYDETLFGSPAGCRPPRPDPDPWVEKA 60  
DB 1 MKTVEVLAVSGMOTLGIQHRRCGGYRYKARISYDETLFGSPAGCRPPRPDPDPWVEKA 60

QY 61 NRTGCVGEAKSALGAKGSCCETTPSRGSPITLTPRKKNKYPISHTSPCYCESLFGSRSR 120  
DB 61 NRTGCVGEAKSALGAKGSCCETTPSRGSPITLTPRKKNKYPISHTSPCYCESLFGSRSR 120

QY 121 GASFGAPPMKAGDAKLRALLMTPTPTPRGSHSPRPAPL 161  
DB 121 GASFGAPPMKAGDAKLRALLMTPTPTPTPRGSHSPRPAPL 161

QY 121 GASFGAPPMKAGDAKLRALLMTPTPTPTPRGSHSPRPAPL 161  
DB 121 GASFGAPPMKAGDAKLRALLMTPTPTPTPRGSHSPRPAPL 161

RESULT 4  
Q08S148 PRELIMINARY; PRT; 647 AA.

AC 08S148;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Vasa1-like protein.  
GN Name=P0042A10.17;  
OS Oryza sativa (Japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzoae; Oryza.  
OX NCBI\_TaxID=39947;  
[1]  
RP SEQUENCE FROM N.A.  
RX PubMed=12447438; DOI=10.1038/nature01184;  
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,  
RA Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
RA Hoshikawa S., Maebayashi M., Arikawa K., Chiden Y., Hayashi M.,  
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
RA Ikono M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,  
RA Karasawa W., Katagiri S., Kikuta A., Kodayashi N., Kono T.,  
RA Machida K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,  
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
RA Namiki N., Negishi M., Ohta I., Ono N., Saito S., Sakai K., Shibata M.,  
RA Shimokawa T., Shomura A., Song Y., Takazaki Y., Terasawa K., Tsuji K.,  
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,  
RA Zhong H., Iwama H., Endo T., Ito H., Hainu J.H., Kim H.I., Eun W.Y.,

RA Yano M., Jiang J., Gojobori T.;  
 RT "The genome sequence and structure of rice chromosome 1.";  
 RL Nature 420:312-316(2002).  
 DR EMBL, AF003343; BAB90075.1; -.  
 RC Gramene; O88148; -.  
 SO SEQUENCE 647 AA; 62062 MW; C27F3FA0A77D1EC3 CRC64;

Query Match 10.4%; Score 150; DB 2; Length 647;  
 Best Local Similarity 26.7%; Pred. No. 0.17; Mismatches 113; Indels 64; Gaps 13;  
 Matches 72; Conservative 21;

QY 23 GGYVAKARTSYVDLTFGSPAGTRPTPPDPP-PWEKANTRGVGEAKALGAKSGCE 81  
 DB 59 GGYNPSPSISGTSPTTGGGGGYTPPSPDTPSPSSDTPSPSGCSSPT-----PCD 113  
 QY 82 TTPRGS-TPTLTPRKKNKXRPISHTPSYCDLSFGSRSEASGAPRMKADAKRAL 140  
 DB 114 APPSPSDTSPPTTGGGGGYSP--TPS-----DTPSPSSD----- 147  
 QY 141 LMTPTPTP--RGSHPREAP-----LRAIHPAG-----PSKTEPCPADSQKLSM 185  
 DB 148 --TSPPTPGGGGGYTPPSDAPSPSSDTPPTTGGGGGYTPPSDAPSPSSDTPPTP 205  
 QY 186 GGLHSSPLKRGSLSHLNLNVPSTGHPATSAPTTNGPDLRPST-----GVTFR 236  
 DB 206 GGGGGYTPPSDTPSPSSSGSSPTT--PGGGGYTPPTPSPSPSSGSSRTTTPGCGCTP 263  
 QY 237 SPLVTSRARSYSIVSPSTRRGATOKPKP 266  
 DB 264 TPCGTPPAPSSGTS-PTTP--GGSYTPPTP 290

## RESULT 5

Q8MQG8 ID Q8MQG8 PRELIMINARY; PRT; 518 AA.

AC Q8MQG8  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Prion-like-(Q/n-rich)-domain-bearing protein protein 75, isoform b.  
 GN Name=pqn-75; ORFNames=W03D2.1;  
 OS Caenorhabditis elegans;  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RG Wormbase Consortium;  
 RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Rohlfing T., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid W03D2.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN 13  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 RN 14  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilson R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN 15  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;

RA Wilson R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN 16  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RG Wormbase Consortium;  
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF000298; AA097860.1; -.  
 DR Wormbase; WBGene00004157; W03D2.1.  
 DR Wormpep; W03D2.1b; CB31729.  
 SO SEQUENCE 518 AA; 50473 MW; 4801F27D9663651E CRC64;

Query Match 10.1%; Score 144.5; DB 2; Length 518;  
 Best Local Similarity 24.1%; Pred. No. 0.3; Mismatches 97; Indels 121; Gaps 13;  
 Matches 75; Conservative 18;

QY 13 QTLGLHRCRGYVAKARTSYVDLTFGSPAGTRPTPPDPP-PWEKANTRGVGEAKK 72  
 DB 213 QGFQQQQQTNQNGFRGRQ-----APPAGSPPPP--PP-----KQSP 247  
 QY 73 ALGAKGSCETTPSRGSTPTLTPRKKNKXRPISHTPSYCDLSFGSRSEASGAPPMAYG 132  
 DB 248 PLAGSGSPPPPPAAGSP--PPRTGSPPPPTGSP-----PPPPAGG 287  
 QY 133 DAAKLALMTPTPTPGS-----HSRRP-----REAPLRAIHPA 167  
 DB 288 SPPTPAGSPPP 347  
 QY 168 GPSTKTP-----GPAADSQLS-----MGQ-----LHSSRLKRL 198  
 DB 348 SPPTGSPPTGPRPGGPGKSSSESSSREGPRGCGRGGKSSSESSSREPRGP 407  
 QY 199 SHSLHTLNPSTGHPATSAPTTNGPDLRPSTGCTFRPLVTSRARSYSIVSPTRRG 258  
 DB 408 RRS-----PPTGSPPTGPRPGSPPTG-----SPTGLPGRQKQ 447  
 QY 259 GATQKP--KPP 267  
 DB 448 APEDRPTGSP 458

## RESULT 6

O02123 ID O02123 PRELIMINARY; PRT; 524 AA.

AC O02123  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Prion-like-(Q/n-rich)-domain-bearing protein protein 75, isoform a.  
 GN Name=pqn-75; ORFNames=W03D2.1;  
 OS Caenorhabditis elegans;  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RG Wormbase Consortium;  
 RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Rohlfing T., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid W03D2.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN 13  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;

RA Waterston R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilson R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilson R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RG WormBase Consortium;  
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF000298; AAC48255.2;  
 DR WormBase; WBGene0004157; W03D2.1.  
 DR WormPep; W03D2.1a; CE26501.  
 KM Prion.  
 SQ SEQUENCE 524 AA; 51210 MW; 023D77EFD172CC5 CRC64;  
 Query Match 10.1%; Score 144.5; DB 2; Length 524;  
 Best Local Similarity 24.1%; Pred. No. 0.3;  
 Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;  
 QY 13 QTLGLOHRCRGYKARTSYVDLTFGSPAGTPTPPDPVVEKANKRTRGVGEKASK 72  
 DB 219 QGFGQOQOTQNGFRGKQ-----APGASPPPP-----PP-----KQSP 253  
 QY 73 ALGAKSGCETTPSGSTPTLTPRKKNKRYPISHTPSYCDESLFGSRSGASFGAPRMKG 132  
 DB 254 PLASGSGPPPPAAGSP--PRTGSPPPPTGSP-----PPPPAGG 293  
 QY 133 DAAKLRLALWTPPTPRGS-----HSRP-----REAPLRAIHPA 167  
 DB 294 SPPPRAGSPPPPPRGSPPTGSLPPQAGSPPPAAGTSPPPPRQKQAPERSPTG 353  
 QY 168 GPSKTEP-----GPAADSQKLS-----MCG-----LHSSRPLKXGL 198  
 DB 354 SPTGSPPTGPRPGPKSSESESESRGCGPRGCGPRGPKRSSESESESESRGCGP 413  
 QY 199 SHSLHLNVSTGHPATSAHPHNGQDLRPTSGVTFSPPLVTSARASVTSVSTPRRG 258  
 DB 414 RRS-----PPTGSPPTGSPPTGPRGSPPTG-----SPTGLPSRQKQ 453  
 QY 259 GATOKP--KPP 267  
 DB 454 APEDRPTGSP 464

RESULT 7  
 Q8WOG9 PRELIMINARY; PRT; 539 AA.  
 ID Q8WOG9  
 AC Q8WOG9  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Prion-like-(Q/n-rich)-domain-bearing protein 75, isoform C.  
 GN Name=pqn-75; ORFNames=W03D2.1;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peleoderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINB=99069613; PubMed=9851916;  
 RG WormBase Consortium;  
 RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Rohlfing T., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid W03D2.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilson R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilson R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RG WormBase Consortium;  
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF000298; AAM97961.1;  
 DR PIR; T28770; T28770.  
 DR WormBase; WBGene0004157; W03D2.1.  
 DR WormPep; W03D2.1c; CE14506.  
 KM Prion.  
 SQ SEQUENCE 539 AA; 52555 MW; D3D3CFD8EF42CDB9 CRC64;  
 Query Match 10.1%; Score 144.5; DB 2; Length 539;  
 Best Local Similarity 24.1%; Pred. No. 0.31;  
 Matches 75; Conservative 18; Mismatches 97; Indels 121; Gaps 13;  
 QY 13 QTLGLOHRCRGYKARTSYVDLTFGSPAGTPTPPDPVVEKANKRTRGVGEKASK 72  
 DB 214 QGFGQOQOTQNGFRGKQ-----APGASPPPP-----PP-----KQSP 268  
 QY 73 ALGAKSGCETTPSGSTPTLTPRKKNKRYPISHTPSYCDESLFGSRSGASFGAPRMKG 132  
 DB 269 PLASGSGPPPPAAGSP--PRTGSPPPPTGSP-----PPPPAGG 308  
 QY 133 DAAKLRLALWTPPTPRGS-----HSRP-----REAPLRAIHPA 167  
 DB 309 SPPPRAGSPPPPPRGSPPTGSLPPQAGSPPPAAGTSPPPPRQKQAPERSPTG 368  
 QY 168 GPSKTEP-----GPAADSQKLS-----MCG-----LHSSRPLKXGL 198  
 DB 369 SPTGSPPTGPRPGPKSSESESESRGCGPRGCGPRGPKRSSESESESESRGCGP 428  
 QY 199 SHSLHLNVSTGHPATSAHPHNGQDLRPTSGVTFSPPLVTSARASVTSVSTPRRG 258  
 DB 429 RRS-----PPTGSPPTGSPPTGPRGSPPTG-----SPTGLPSRQKQ 468  
 QY 259 GATOKP--KPP 267  
 DB 469 APEDRPTGSP 479

RESULT 8  
 O6NVE5 PRELIMINARY; PRT; 1067 AA.  
 ID O6NVE5  
 AC O6NVE5;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein LOC73072.  
 GN Name=LOC73072;  
 OS Mus musculus (Mouse).



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L.H., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien P.,  
 RA Diatchenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Brownstein M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywiński M.I., Skalska U., Smalins D.E., Schmeich A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Straubeberg R.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC068157; AA68157.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1067 AA; 107615 MW; 561B9289E89FE824 CRC64;  
 Query Match 10.0%; Score 144; DB 2; Length 1067;  
 Best Local Similarity 26.7%; Pred. No. 0.66;  
 Matches 74; Conservative 25; Mismatches 118; Indels 60; Gaps 13;  
 QY 37 TLEGSP---AGTPTPPDFPPVWEKANR-----TRGVGKASALGAKGSCETTP---SR 86  
 DB 106 TTAGAPRRSAGAPRRPAPRRPPASGKERAPMKTPGGGSISSGRRASGIGARBPVQKR 165  
 QY 87 GSTPTLTP---RKXKXKRPISHTPSYCDLSFGSRSEGSFGAP-----RWAKDAKRLA 139  
 DB 166 LQPTPEPSARKGTPETPKNTLNSGRRRLVSDLSLGTSGAPSPATIRSRAPATB--- 222  
 QY 140 LMTPTPTPGSHSPRRPRAIPAIHPAG--PSKTEPGPADSGLKSLMGSLHSSRPPLKRL 198  
 DB 223 -VGLPQAPAPARORPLTEAARPKGSSASRPSATELSPAFRRRSVAGGSL--QKPSRSL 279  
 QY 199 SHSLTLHLNVSTG-----HFA--TSAHTTNP--QDLRPSI 230  
 DB 280 IBSATPOLSPBSRGVSPRTVPRAPAHTSQLSKGGQALHPTQTVPRKKNPSVOSLIPAS 339  
 QY 231 SGVTFPSPLVTSRRASVSISVPSPTPRGATQKRP 267  
 DB 340 SLVTPTPP-----GASSVGGPDDPSQ--TTLRPSFP 368  
 RESULT 9  
 Q26963 PRELIMINARY; PRT; 964 AA.  
 AC Q26963;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DE Trans-sialidase.  
 DE Name=CTS-121;  
 OS Eukaryota; Euteleostomi; Euteleostomi; Euteleostomi;  
 OC Eukaryota; Euteleostomi; Euteleostomi; Euteleostomi;  
 OC NCBI\_TaxID=5693;  
 RN [1]

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=93010978; PubMed=1396577;  
 RA Uemura H., Schenkman S., Nussenzweig V., Eichinger D.;  
 RT "Only some members of a gene family in Trypanosoma cruzi encode  
 RT proteins that express both trans-sialidase and neuraminidase  
 RT activities.";  
 RL EMBO J. 11:3837-3844(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=97001677; PubMed=844669; DOI=10.1016/0166-6851(96)02634-5;  
 RA Smith L.E., Uemura H., Eichinger D.;  
 RT "Isolation and expression of an open reading frame encoding sialidase  
 RT from Trypanosoma fangeli.";  
 RL Mol. Biochem. Parasitol. 79:21-33(1996).  
 DR HSP; Q26964; IMS3.  
 DR GO: GO:0004308; F:exo-alpha-sialidase activity; IEA.  
 DR GO: GO:0009405; P:pathogenesis; IEA.  
 DR InterPro: IPR008985; ConA\_like\_1ec\_g1.  
 DR InterPro: IPR002860; Glyco\_hydro\_BNR.  
 DR InterPro: IPR008377; Sialidase\_trypan.  
 DR Pfam: PF02012; BNR; 3.  
 DR PRINTS: PRO1803; TCSIALIDASE.  
 SQ SEQUENCE 964 AA; 101610 MW; 8F0600289FB847AD CRC64;  
 Query Match 9.9%; Score 142.5; DB 2; Length 964;  
 Best Local Similarity 26.8%; Pred. No. 0.74;  
 Matches 84; Conservative 24; Mismatches 109; Indels 97; Gaps 19;  
 QY 13 QTLGLHRCRGGR-----VKARTSYDELTLFGSPA----- 43  
 DB 609 RTPDISHFYVGGRSDMPTISHVTNNVLLYNNROLMAEIRTLFLSODLIGTEAMDS 668  
 QY 44 -----GTAPTPPDPPVWEKANRTRGV-----GKEASALGAKGSCETTPS-----RG 87  
 DB 669 SDSNAHSTSTPD-----SSAHSTPPADNGAHSTSTPADNGAHSTSTPDNGAH 722  
 QY 88 STPTLTPRRKKNRPISHTPSYCDLSFGSRSEGSFGAPMAKGDADALRLMTPTPT 147  
 DB 723 STP-LTPADNGAH---STPTPADSNAHSTSTPADNGA-----HSTPT 763  
 QY 148 P--RGSHPRRPRAIPAIHPA--GSKTEPGPADSGLKSLMGSLHSSRPPLKRL 195  
 DB 764 PADNGAHS-----TPTL--PADNGAHSSTPTP-ADSSAHSTPPADNGAHSTSTPAD 814  
 QY 196 RGLSHSLTLHLNVSTGHPATSPATHTNGPOD--LRPSTGVTFRSPVTSRRASVSISPS 253  
 DB 815 NG-AHSTPTPADSNAHSTSTPPADNGAHSTSTPADNGA-HSTPL--TPADSNAHSTPS 870  
 QY 254 TPRRGATQKRP 267  
 DB 871 TPGDNGAHSTPSAP 884  
 RESULT 10  
 SYNI CANFA STANDARD; PRT; 415 AA.  
 ID SYNI CANFA  
 AC 062732;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Synapsin-1 (Synapsin I) (Fragment).  
 GN Name=SYNI;  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OC NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Shibuya H., Liu P.-C., O'Brien D.P., Johnson G.S.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,  
 binds to the cytoskeleton, and is believed to function in the  
 regulation of neurotransmitter release (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Synapse.  
 CC -1- SIMILARITY: Belongs to the synapsin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF049588; AAC05207.1; -.  
 DR HSSP; P17599; 1AUX.  
 DR InterPro; IPR001359; Synapsin.  
 DR Pfam; PF02750; Synapsin C; 1.  
 DR PRINTS; PR01368; SYNAPSN.  
 DR PROSITE; PS00415; SYNAPSN\_1; PARTIAL.  
 DR PROSITE; PS00416; SYNAPSN\_2; PARTIAL.  
 KW Actin-binding; Phosphorylation; Repeat; Synapse.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 148 C (actin-binding and synaptic-vesicle  
 binding).  
 FT DOMAIN 149 383 D (Pro-rich linker).  
 FT DOMAIN 384 >415 E.  
 FT NON\_TER 415 415  
 SQ SEQUENCE 415 AA; 43388 MW; E988E2026FC5361B CRC64;

Query Match 9.7%; Score 140; DB 1; Length 415;  
 Best Local Similarity 26.0%; Pred. No. 0.45;  
 Matches 63; Conservative 17; Mismatches 108; Indels 54; Gaps 9;

QY 40 GSPAGTRPPDPDPVVEKANKTRG-----VGRKASKALGAKSCETTPRSSTP 90  
 DB 182 GPFAQQR-PPQGGPFGPGPQKQKQPPPLQQRPPPGQGHISGLSPGPG-SPLPRLSP 239  
 QY 91 TLPRKKNKYPPISTHTPSYCDSESLFGSRSEGSFGAPPMKADAKLRLMTPTPTPR- 149  
 DB 240 TSAQ-----QPVSGAQPLSGAQKQSRPVAGGPAPAPAPAPASPSRQKQAPQARQ 294  
 QY 150 -----GSHSPRRAPRLAIHPAGSEKTPPPALDSQKLSMGLSHSRPLKGLSHSLTHL 205  
 DB 295 TSVSGAAPPKASGAPPSGQKQKPPGPPGAPTRQASQAG----- 336  
 QY 206 NVPSGTGHPATAPHTNGPQDLRPSSTGVTFRSPVLTSSARSVSISSPTPRRGATQKPK 265  
 DB 337 PMPRTGPTTQOP-----RPSGPPAGR-PAKPKQLAKRSPDVPSPATAGG----- 381  
 QY 266 PP 267  
 DB 382 PP 383

RESULT 11  
 ID 06C708 PRELIMINARY; PRT; 659 AA.  
 AC 06C708;  
 DT 25-OCT-2004 (TREMBLrel. 28, Created)  
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)  
 DB Similar to tr|Q95JC9 Sus scrofa Basic proline-rich protein.  
 GN ORFNames=YALI0D261919;  
 OS Yarrowia lipolytica CLIB99.  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 RX NCB1\_TaxID=284591;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIB99;

RG Genolevures;  
 RA Dulton B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neveglie C., Talla E.,  
 RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,  
 RA Baray S., Blanchon S., Beckrich J.M., Beyne E., Bleykasten C.,  
 RA Bolerae A., Boyer J., Catterico L., Confantolero F., de Darvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet M., Groppi A.,  
 RA Hantraye F., Henequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicoud J.M., Nikoliski M., Ozias S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swenne D., Tekla F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zenou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.L.;  
 RT "Genome evolution in yeasts."  
 RN Nature 430:35-44(2004).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIB99;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR382130; CAG81510.1; -.  
 DR InterPro; IPR003124; WH2.  
 DR Pfam; PF02205; WH2; 1.  
 SQ SEQUENCE 659 AA; 62466 MW; 2CE5088BC9EF0908 CRC64;

Query Match 9.6%; Score 138.5; DB 2; Length 659;  
 Best Local Similarity 26.0%; Pred. No. 0.88;  
 Matches 70; Conservative 15; Mismatches 111; Indels 73; Gaps 11;

QY 42 PACGTRPPDPDPVVEKANKTRGVGKASALGAKSCETTPRSSTPILTPRKKNK 101  
 DB 387 PGGAAPPIPGSAAPAPAPAPAGAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 442  
 QY 102 PISHTPSYCDSESLFGSRSEGSFGAPPMKADAKLRLMTPTPTPSGS-----PRP 156  
 DB 443 PPSPEPS-----TPGPMFGAP-MPKPAP-----ASGAPPPPPGAPAPAPAPAPAPAP 490  
 QY 157 REAPRLAIHPAGPSTKTEBPAA-----DSQKL-----SMGLSHSRPLKGLSHSLTHLV 207  
 DB 491 PAQPPSPGAPPPGAPPPGAPPPPTDGFHSMILDDSGSGHGAP-----P 537  
 QY 208 PSTGHPATAPHTNGPQDLRPSSTGVTFRSPV-----TSRARSVS 248  
 DB 538 PPSAPPSNGSHSGAPPPPPNGVNRKRDVORTSTLGSNNIRITLDTSAYTAPRAVS 597  
 QY 249 ISVPTPRRGATQKPKP-----WK 269  
 DB 598 TPVSSSGGGGGS---KPPQIKIDSRWK 623

RESULT 12  
 ID 075MN6 PRELIMINARY; PRT; 2185 AA.  
 AC 075MN6;  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
 DE Hypothetical protein MLL3 (Fragment).  
 GN Name=MLL3;  
 OS Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;  
 RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,  
 RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,  
 RA Wylie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,  
 RA Fewell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,  
 RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Tsak A.,

RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,  
 RA Ozerky P., Bielicki L., Scott K., Holmes A., Harting R., Harris A.,  
 RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,  
 RA Kozlowski-Reilly A., Leonard S., Rohlfing T., Rock S.M.,  
 RA Tin-Wollam A.M., Abbott A., Minx P., Maupin R., Strommatt C.,  
 RA Latteille P., Miller N., Johnson D., Murray J., Woessner J.P.,  
 RA Wendi M.C., Yang S.P., Schultz B.R., Wallis J.W., Spieth J.,  
 RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.B., Cook L.L.,  
 RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,  
 RA Clifton S.W., Chissee S.L., Marra M.A., Raymond C., Haugen E.,  
 RA Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubb K.,  
 RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Pury T.S.,  
 RA Baertsch R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,  
 RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,  
 RA Bddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,  
 RA Waterston R.H., Wilson R.K.,  
 RT "The DNA sequence of human chromosome 7.",  
 RL Nature 424:157-164(2003).

RT [2]  
 RN

RP SEQUENCE FROM N.A.  
 RA Bauer C., Gaije T., Nguyen C.,  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.

RA Wilson R.,  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC006474; AA000364.1; -  
 DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000194; ATPase a/b-centre.  
 DR InterPro; IPR011011; FYVE\_PHD\_ZnF.

DR InterPro; IPR000910; HMG\_12\_box.  
 DR InterPro; IPR001965; ZnF\_PHD.

DR Pfam; PF00505; HMG\_box; 1.  
 DR Pfam; PF00628; PHD\_3.

DR SMART; SM00398; HMG; 1.  
 DR SMART; SM00249; PHD; 3.

DR PROSITE; PS00152; ATPASE ALPHA BETA; UNKNOWN\_1.  
 DR PROSITE; PS01359; ZF PHD 1; UNKNOWN\_1.

DR PROSITE; PS50016; ZF PHD 2; 1.  
 DR Hypothetical protein\_

KM NON TER 2185 2185  
 FT SEQUENCE 2185 AA; 239412 MW; 73A98266F33817E CRC64;

SO

Query Match 9.5%; Score 136.5; DB 2; Length 2185;  
 Best Local Similarity 24.0%; Pred. No. 4;

Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARSYVDDELFGSPAGTR-----PTPPDFDPWVEK-----ANRTGVGK 68

DB 994 SRLQNNETVANNPSEVRLCSSSTNNNDPYAKPPPTPPVMTDQPKSLGLSRSPVSE 1053

QY 69 EASKALGAKGSC-----ETTPSGSGPTLTLPKKN-----YRPSHTPS 108

DB 1054 QIRKGIAGTSHFTKPSPPRADVFOROKI PDSYARPLTPADLDSGPGFPTPMOPPS 1113

QY 109 YCDSELFSGRSFGASFGAPRMAGDAKRLALMTPTPTPGSH--SPRPREAPLRAIP 166

DB 1114 SQDP--YGSVSG-----ASRLSLVDYERPAL--TRPILDNSSHNSNDPYSPGLTTRIP 1164

QY 167 -----AGSK--TEPG-----PAADSQKLSNGILHS-----SRPL 194

DB 1165 AVNESFAHSPRAFGPGTISRPTSDPYSPGPTPPVVDYSQSOSGSTARSTNDPYSGP- 1223

QY 195 KRGLSHSLTHLWNPSTGHPATSAPHNNGPDLRPSVGTTPSPSLVTSARASVTSVPT 254

DB 1224 -----PGTPPTTVDPYSGQQTPTPTSTQTLFTVPTVNHSHSDPYAHPET 1270

QY 255 PRGGATQKPKPP 267

DB 1271 PRGGSIVPYSGPP 1283

RESULT 13

ML3\_HUMAN STANDARD; PRT; 4911 AA.

ID Q9NE24; Q9NC02; Q9NDF6; Q9H9P4; Q9NR13; Q9P222; Q9UDR7;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-lysine N-methyltransferase, H3 lysine-4 specific ML3) (EC 2.1.1.43)

DE (Homologous to ALR protein)

GN Name=ML3; Synonyms=HALR, KIAA1506;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxId=9606;

RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Fetal thymus;  
 RX MEDLINE=21888622; PubMed=11891048; DOI=10.1016/S0378-1119(02)00392-X;

RA Rualit M., Brun M.-E., Ventura M., Roizes G., De Sario A.;  
 RT "ML3, a new human member of the TRX/MLL gene family, maps to 7q36, a  
 chromosome region frequently deleted in myeloid leukaemia.";  
 RL Gene 284:73-81(2002).

RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Cervical carcinoma;  
 RX MEDLINE=21574953; PubMed=11718452;

RA Tan Y.C., Chow V.T.;  
 RT "Novel human HALR (ML3) gene encodes a protein homologous to ALR and  
 RT with leukemia and developmental defects.";  
 RL Cancer Detect. Prev. 25:454-469(2001).

RN [3]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;  
 RA Hallier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,  
 RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,  
 RA Wylie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,  
 RA Fwell G.A., Delaunay K.D., Miner T.L., Nash W.E., Cordes M., Du H.,  
 RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,  
 RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,

RA Ozerky P., Bielicki L., Scott K., Holmes A., Harting R., Harris A.,  
 RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,  
 RA Kozlowski-Reilly A., Leonard S., Rohlfing T., Rock S.M.,  
 RA Tin-Wollam A.M., Abbott A., Minx P., Maupin R., Strommatt C.,  
 RA Latteille P., Miller N., Johnson D., Murray J., Woessner J.P.,  
 RA Wendi M.C., Yang S.P., Schultz B.R., Wallis J.W., Spieth J.,  
 RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,  
 RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,  
 RA Clifton S.W., Chissee S.L., Marra M.A., Raymond C., Haugen E.,  
 RA Gillett W., Zhou Y., James R., Phelps K., Iadamoto S., Bubb K.,  
 RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Pury T.S.,  
 RA Baertsch R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,  
 RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,  
 RA Waddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,  
 RA Waterston R.H., Wilson R.K.,  
 RT "The DNA sequence of human chromosome 7.";  
 RL Nature 424:157-164(2003).

RN [4]  
 RP SEQUENCE OF 556-3865 FROM N.A. (ISOFORM 1).

RC TISSUE=Brain;  
 RX MEDLINE=20277482; PubMed=10819311.

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:143-150(2000).

RN [5]  
 RP SEQUENCE OF 3193-3865 AND 4460-4911 FROM N.A.

RC TISSUE=Placenta;  
 RX PubMed=14702039; DOI=10.1038/ng1285;

RA Ota T., Suzuki Y., Nishikawa T., Ohtsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,

RA Sekinm M, Obayashi M., Nishi T., Shihahara T., Tanaka T., Ishi S.,  
RA Saito K., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Negahari K., Murakami K., Yaenda T., Iwayanagi T., Wagatsuma M.,  
RA Shitatori A., Sudo H., Hosoliti T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
RA Yamazaki M., Ninomiya K., Ieshibshi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimeta M., Watanabe M., Hizoka S., Chiba Y.,  
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houta T.,  
RA Kusanu J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togiya S., Komai F., Hara R., Takuchi K., Arita M.,  
RA Imose N., Mueselino K., Yuki F., Oshima A., Saeki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Teshima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumaigi A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Koniyama M., Tashiro T., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Saeki M.,  
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Matsushima-Sugano J., Saich T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs." ;  
RL Nat. Genet. 36:40-45(2004) .  
RP [6] ;  
RP SEQUENCE OF 3879-4911 FROM N.A.  
RC TISSUE=Testis ;  
RC Duesterhoeft A., Lauber J., Mewes H.-W., Weil B., Wiemann S.,  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBD databases.  
RP [7] ;  
RP INTERACTION WITH ASC-2/NGO6 CONTAINING COMPLEX (ISOPFORM 2) .  
RC TISSUE=Cervical carcinoma; PubMed=12482968; DOI=10.1128/MCB.23.1.140-149.2003;  
RC MEDLINE=22371146; Pubmed=12482968; DOI=10.1128/MCB.23.1.140-149.2003;  
RA Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,  
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,  
RA Acores D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,  
RA Lee J.W. ;  
RT "Activating signal cointegrator 2 belongs to a novel steady-state  
RT complex that contains a subset of trithorax group proteins." ;  
RL Mol. Cell. Biol. 23:140-149(2003) .  
RL -1- FUNCTION: Belongs to the ASC-2/NGO6 complex (ASCOM), a  
CC coactivator complex of nuclear receptors, involved in  
CC transcriptional coactivation. MLL2 may be a catalytic subunit of  
CC this complex, which weakly methylates Lys-4 of histone H3. This is  
CC a specific tag for epigenetic transcriptional activation. May be  
CC involved in leukemogenesis and developmental disorder.  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =  
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.  
CC -1- SUBUNIT: Belongs to the ASC-2/NGO6 complex (ASCOM), which  
CC contains ASC-2/NGO6, the retinoblastoma-binding protein RbQ-3/  
CC RBBP3, alpha- and beta-tubulins, the trithorax group proteins MLL2/  
CC and MLL1, and ASH2/ASCL2. Interacts with histone H3.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=O8NEZ4-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=O8NEZ4-2; Sequence=VSP\_008561, VSP\_008562;  
CC -1- TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed  
CC by brain and liver. Also expressed in placenta, peripheral  
CC blood, fetal thymus, heart, lung and kidney. Within brain,  
CC expression was highest in hippocampus, caduate nucleus, and  
CC substantia nigra. Not detected in skeletal muscle and fetal liver.  
CC -1- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B  
CC and H4, and may have a H3 lysine specific methylation activity.  
CC -1- MISCELLANEOUS: Found in a critical region of chromosome 7, which  
CC is commonly deleted in malignant myeloid disorders. Partial

CC duplication of the MLL3 gene are found in the juxtaacromeric  
CC region of chromosomes 1, 2, 13 and 21. Juxtaacromeric  
CC reshuffling of the MLL3 gene has generated the BAGE genes.  
CC -1- SIMILARITY: Belongs to the TRX/MLL family.  
CC -1- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.  
CC -1- SIMILARITY: Contains 1 DHHC-type zinc finger.  
CC -1- SIMILARITY: Contains 6 PHD-type zinc fingers.  
CC -1- SIMILARITY: Contains 1 post-SET domain.  
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.  
CC -1- SIMILARITY: Contains 1 SET domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL, AY024361, AAK0583.1, -,  
DR EMBL, AF264750, AAE74766.2, -,  
DR EMBL, AC006017, AAD45842.1, -,  
DR EMBL, AC104592, -, NOT\_ANNOTATED\_CDS.  
DR EMBL, AC005631, -, NOT\_ANNOTATED\_CDS.  
DR EMBL, AB040939, BA96030.2, -,  
DR EMBL, AK022687, BAB14179.1, -,  
DR EMBL, AK075113, BAC11409.1, -,  
DR EMBL, AB833924, CAD38780.1, -,  
DR HSSP, Q14839, 1MM2.  
DR Genew, HGNC:13726, MLL3.  
DR MIM, 606833, -,  
DR InterPro, IPR000637, A+T\_hook.  
DR InterPro, IPR003889, Fyrich\_C.  
DR InterPro, IPR003888, Fyrich\_N.  
DR InterPro, IPR011011, FYVE\_PHD\_ZnF.  
DR InterPro, IPR000910, HMG\_12\_box.  
DR InterPro, IPR003616, PostSET.  
DR InterPro, IPR001214, SET.  
DR InterPro, IPR001594, Znf\_DHHC.  
DR InterPro, IPR001965, Znf\_PHD.  
DR InterPro, IPR001841, Znf\_ring.  
DR Pfam, PR02178, AT\_hook.1.  
DR Pfam, PF05965, FYR1.  
DR Pfam, PF05964, FYRN.1.  
DR Pfam, PF00505, HMG\_box.1.  
DR Pfam, PF00628, PHD\_6.  
DR Pfam, PF00856, SET.1.  
DR PROSITE, PS00354, HMG\_Y.1.  
DR PROSITE, PS0868, POST\_SET.1.  
DR PROSITE, PS50280, SET.1.  
DR PROSITE, PS50216, ZF\_DHHC.1.  
DR PROSITE, PS01359, ZF\_PHD\_1.5.  
DR PROSITE, PS50016, ZF\_PHD\_2.6.  
DR PROSITE, PS50089, ZF\_RING\_2.1.  
KW Activator; Alternative splicing; Chromatin regulator; Coiled coil;  
DNA-binding; Methylyltransferase; Nuclear protein; Polymorphism; Repeat;

	Query Match	9.5%;	Score 136.5;	DB 1;	Length 4911;
	Best Local Similarity	24.0%;	Pred. No. 9;		
	Matches	75;	Conservative	28;	Mismatches 113; Gaps 13;
QY	29	AFTSYVDLTFGSPACR-----	PTPPDDEPWBEK-----	-AARTGYCK	68
Db	1933	SRLQNNETTANRPSPVRLCSSSTNNNDYAKPPTPRPVMTDQPKSIGLSRSVSSE			1992
QY	69	EASKALGAGKSC-----	ETTPSKSGPTLLPRKNK----	-YRPSHTPS	108
Db	1993	QTAKGPITAGTSDHFTFKSPRADVFQRORI	PDYSAPPLLPADLDGPGCFKTPWPPPS		2052
QY	109	YCDESILFGSRSECAFGAPMAKGDAKLRIALLMTPPTPRGSH--	SPPREAPLRAIHP		166
Db	2053	SDDP---YGAVSQ-----	ASRRSLVDYEERPAL--TRRPIDNFSHNSQNDPYSOPIPLPHD		2103

```

QY 167 -----AGPSK--TEPG-----PAADSOXKLSMGLHS-----SRPL 194
DB 2104 AVNESFAHPSRAFSQPGTISNPTSDQDPYSGPPGTRRVVDSYSSGSSSTASNTDPPYSGP- 2162
QY 195 KRGLSHSLTHLNVSTGHFATSAHTNGPODLRPSIGGVTFRSPVLVTSRAVSISVSPST 254
DB 2163 -----GTRPRTVDPVYSQOPQTFRPSQTQDLVFTVTVNQHRSHDPYAHPPGT 2209
QY 255 PRRGATOKPKPP 267
DB 2210 PRRGISVPSQPP 2222

RESULT 14
Q9LJL1 PRELIMINARY; PRT; 769 AA.
AC Q9LJL1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DE Calx protein.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20379056; PubMed=10807915; DOI=10.1074/jbc.M002686200;
RA Pan J., Snell W.J.;
RT "Regulated targeting of a protein kinase into an intact flagellum. An
RT autoara/flip-like protein kinase translocates from the cell body into
RT the flagella during gamete activation in chlamydomonas."
RL J. Biol. Chem. 275:24106-24114(2000).
CC -1- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR HSPB; AF199021; AAF9501.1; -
DR HSPB; AF199021; AAF9501.1; -
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR011009; Kinase like.
DR InterPro: IPR002290; Ser. Thr. kinase.
DR InterPro: IPR008271; Ser. Thr. pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD000001; Prot. Kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 769 AA; 80249 MW; DB037FFCD4245101 CRC64;

Query Match 9.5%; Score 136; DB 2; Length 769;
Best Local Similarity 22.8%; Pred. No.1.5; Indels 98; Gaps 15;
Matches 75; Conservative 41; Mismatches 15;

QY 32 SYVDETLFG-----SPAGTRTPPPDDPP-----WVERANR----- 62
DB 323 SEVEPQQLGFRGSGNAGCGSPSGSKPPRLQVSSVAGGRLGHSSMSAMDKRYRD 382
QY 63 -----TRGVKKAASKAL-GAKSCCTTSRSGSTPLTLTRKKKK-----YRPIH 105
DB 383 HILDVTLGMAAGAAATLALKSVKVSIGDGTGGPSSAAGARVSRAGTANNSGTGAFLSR 442
QY 106 TRSYCDSESLFGSRSE-----GASFGAPRMAKGAALRALILW-----TPPTPPGS 151
DB 443 NMSMSQAVVAAAAAAMASGSRGAAHSGSDPKGRALTRBELTASLTAPLAGRGSFVL 502
QY 152 -----HSPRPR-APLRAIHPACPSKTERGPADSOQLSMGGLHSSRPLKRG-----LS 199
DB 503 AASFAGRASPAPOQVAPL-----PTWAGKSGGGGGLGSRSSLSGQATAPSPRGAELA 558
QY 200 HSLTHLNVSTGH-----PATSA-----PHTNG-----PDLRPSISVGTFRSP 238

```

```

DB 559 HDLSRLRVDSAAAGCGGAAGPSTISASGAVPYSKSGSSGIRPSPPQAPSPTAGLRTGSD 618
QY 239 LVTSRASVSISVSTPRRGATOKPKPP 267
DB 619 LIPSRL-SQSFSPVSLAAGSNASAGTP 646

RESULT 15
Q7KSE8 PRELIMINARY; PRT; 2556 AA.
AC Q7KSE8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, last annotation update)
DE CG7467-PC.
GN Name=osa; ORFNames=CG7467;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazel R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Gabor G.L.,
RA Abail J.F., Abayavani A., An H.J., Andrews-Pfankoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktiroglu L., Bessley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck Y., Broxstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Waesatman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Ceiniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan M., Frise E., Hodson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svitskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila

```

RT	melanogaster euchromatic genome sequence."					
RL	Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=22426070; PubMed=12537573;					
RA	Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R., Patel S., Frisoe E., Wheeler D.A., Lewis S.E., Rubin G.M., Ashburner M., Celisner S.E.,"The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective."					
RL	Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).					
RN	[4]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=22426069; Pubmed=12537572;					
RA	Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E., Smith C.D., Tsip J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P., Beutenhout B.R., Celisner S.E., de Grey A.D., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;"Annotation of the Drosophila melanogaster euchromatic genome: a systematic review."					
RL	Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).					
RN	[5]					
RP	SEQUENCE FROM N.A.					
RG	FLYBase:					
RL	Submitted (SEP-2002) to the EMBL/genbank/DBJ databases.					
RN	[6]					
RP	SEQUENCE FROM N.A.					
RG	FLYBase:					
RL	Submitted (MAR-2004) to the EMBL/genbank/DBJ databases.					
DR	EMBL; AE003718; AA6516.1; .					
DR	GO; GO:0005622; C:intracellular; IEA.					
DR	GO; GO:003677; F:DNA binding; IEA.					
DR	InterPro; IPR001606; ARID.					
DR	Pfam; PF01388; ARID. 1.					
SO	SMART; SM00501; BRIGHT. 1.					
	SEQUENCE 2556 AA; 268968 MW; 7F0BE49CCB81D21E CRC64;					
	Query Match 9.5%; Score 136; DB 2; Length 2556;					
	Best Local Similarity 22.5%; Pred. No. 5;					
	Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11					
OY	23 GGYRV-----KARTSYDETLFGSPAGTRPTPDF-----DPWEKANTRGVG 67					
Dd	:::					
	633 GGYYKGCGQSGSAGAGYPPOQYPPGVNPVPPQYPPCAVATGPBPPTSQA----GAG 688					
OY	68 KEASALIAFKAKSCETTPRSGTPTPLTPRKKNKYRRISHTPSICDSLSFGSRBEAGSFAP 127					
Dd	669 GANSPSPAQAGC-GYPRGK-----MPNHTGGYPPYQWVPSPPQQTVPGAGAGAMVGNN 741					
OY	128 RMAKDAAKLALALWTLP-----DPTPGSHSPPREARLPAIHPAGSKTEPGPADS 180					
Dd	742 VQGKG-----TPRPVVGGPRPPQGGSSPPRLVYLKHLGHKGGYGSSPTPPQCP 791					
OY	181 QKLSTNG--GLHSRPL-----KRGLSHSLTHLVNST-----GHATSAPHT 220					
Dd	792 QGYNGPPTGMHPRGMGPPHHNGPRPHGPTNNMGRPTSTPPQSOMLQGGQPQGGAGGAPES 851					
OY	221 NGPDLLRPSTGCVTRSPVLVTSRAVSYSIVSTP 255					
Dd	852 GGEPIIT-SQDNGISSGGPGAAGMHAIVTSVVTTGP 885					

```
Search completed: March 28, 2005, 08:53:15
Job time : 87 secs
```

**This Page Blank (uspto)**